Context Free Grammars

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 α_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).

Chomsky Normal Form

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

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

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

Probabilistic Context Free Grammars

A PCFG assigns each rule $X \rightarrow \beta$ a probability $\mathrm{P}(\mathrm{X} \rightarrow \beta)$ satisfying the following.

$$\sum_{\beta} \mathrm{P}(\mathrm{X} \rightarrow \beta) = 1$$

A PCFG assigns a probability to each string.

The Viterbi Algorithm for PCFGs

Consider a PCFG in Chomsky normal form. Considr a string $w_1 \ldots w_n$. We are interested in finding the most probable derivation tree of the given string under th givenb PCFG. Let $\mathrm{Viterbi[X, i, j]}$ be the maximum probability of any single derivations of the string $w_i \ldots w_{j-1}$ form the nonterminal X.

$$\mathrm{Viterbi[X, i, i+1]} = \mathrm{P\left(X \rightarrow w_i\right)}$$

$$j > i+1, \mathrm{Viterbi[X, i, j]} = \max_{k:\ i<k<j,\ X \rightarrow YZ}$$

$$\mathrm{P\left(X \rightarrow YZ\right) Viterbi(Y, i, k) Viterbi[Z, k, j]}$$

Inside Procedure

$$\text{Inside[X, i, j]} = P(X \rightarrow w_i, \ldots, w_{j-1})$$

$$\text{Inside[X, i, i + 1]} = P(X \rightarrow w_i)$$

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

Outside Procedure

$$\text{Outside[X, i, j]} = P(S \rightarrow w_1, \ldots, w_{i-1}, X, w_j, \ldots, w_n)$$

$$\begin{aligned}
\text{Outside[S, 1, n]} &= 1 \\
\text{Outside[Y, i, j]} &= \sum_{X \rightarrow YZ, k > j} \\
&\quad P(X \rightarrow YZ)\text{Outside[X, i, k]Inside[Z, j, k]} \\
&\quad + \sum_{X \rightarrow ZY, k < i} \\
&\quad P(X \rightarrow ZY)\text{Outside[X, k, j]Inside[Z, k, i]}
\end{aligned}$$

Phrase Probabilities

$$P(\mathrm{Phrase(X, i, j)} \mid \mathrm{w_1, \ldots, w_n}) = \frac{\mathrm{Outside[X, i, j] Inside[X, i, j]}}{P(w_1, \ldots, w_n)}$$

$$P(w_1, \ldots, w_n) = \mathrm{Inside[S, 1, n]}$$

Assignment

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