Probabilistic context-free grammars

A. Overview

1. CFGs
2. PCFGs
3. Sentence probabilities

B. What is a CFG?

S → NP VP
NP → V
VP → V NP
V → sees
V → helps
NP → Mindy
NP → Mary

C. What is a PCFG?

A PCFG is a context-free grammar where each rule has an associated probability. In addition, the rules that expand any particular non-terminal A must exhibit a probability distribution, i.e. their probabilities must sum to one (Suppes 1970).
D. **What is a the probability of a sentence?**

\[(6) \quad p(s) = \sum_j p(t_j)p(s|t_j)\]

\[(7) \quad \begin{align*}
\text{NP} & \rightarrow \text{NP C NP .4} \\
\text{NP} & \rightarrow \text{Mary .3} \\
\text{NP} & \rightarrow \text{Mindy .2} \\
\text{NP} & \rightarrow \text{Mark .1} \\
\text{C} & \rightarrow \text{and 1}
\end{align*}\]

\[(8) \quad \text{The probability of each parse is: } .3 \times .2 \times .1 \times 1 \times 1 \times .4 \times .4 = .00096. \text{ The overall probability of the string is } .00096 + .00096 = .00192.\]

\[(9) \quad \begin{array}{c}
\text{NP} \\
\text{NP} \quad \text{C} \quad \text{NP} \\
\text{Mary} \quad \text{and} \quad \text{NP} \quad \text{C} \quad \text{NP} \\
\quad \\
\text{Mindy} \quad \text{and} \quad \text{Mark}
\end{array}\]

\[(10) \quad \begin{array}{c}
\text{NP} \\
\text{NP} \quad \text{C} \quad \text{NP} \\
\text{NP} \quad \text{C} \quad \text{NP} \\
\text{Mary} \quad \text{and} \quad \text{Mindy}
\end{array}\]

**References**

