Lecture 16

NORMAL FORMS OF CFGs

Outline:

- Elimination of useless symbols
- " " ε-productions
- " " chain productions
- Converting CFGs to their Chomsky Normal Forms.

Elimination of useless symbols

Useless symbols are the ones that are not used in any terminal derivation.

Let $G = (V, \Sigma, R, S)$ be a CFG.

A grammar symbol $X \in \Sigma \cup V$ is useful if there is a derivation $S \Rightarrow^* \alpha X \beta \Rightarrow^* w$ where $\alpha, \beta \in (\Sigma \cup V)^*$ and $w \in \Sigma^*$.

Equivalently, X is useful iff it appears in a terminal derivation tree.



v is a substring of w.

Remark: If X is not useful, then it is useless.

In a derivation like $S \Rightarrow^* \alpha X \beta \Rightarrow^* w$, consider the following two subderivations:

- $S \Rightarrow^* \alpha X \beta$ $\quad \alpha, \beta \in (\Sigma \cup V)^*$
- $X \Rightarrow^* v$ $\quad v \in \Sigma^*$

Definition: A nonterminal $X \in V$ is productive if there is a derivation $X \Rightarrow^* v$, $v \in \Sigma^*$.

e.g. $S \to AB$
$B \to bB \mid A$
$A \to a$

$A \Rightarrow a$.

A is productive

$S \Rightarrow AB \Rightarrow aB \Rightarrow aA \Rightarrow aa$.

Definition: A grammar symbol $A \in \Sigma \cup V$ is reachable if there is a derivation $S \Rightarrow^* \alpha A \beta$, $\alpha, \beta \in (\Sigma \cup V)^*$.

Proposition: Let $G = (V, \Sigma, R, S)$ be a CFG. Then all symbols in $(V \cup \Sigma)$ is useful iff

i) Every nonterminal is productive

ii) Every symbol is reachable.

Elimination nonproductive nonterminals:

Idea: First find all the productive nonterminals.

Algorithm (Compute productive nonterminals)

Input: A CFG $G = (V, \Sigma, R, S)$

Output: $P \subseteq V$, set of prod. nonterminals.

$$P_1 := \{ A \in V \mid \exists A \to v \in R : v \in \Sigma^* \}$$

$i = 1$

repeat

$i = i + 1$

$$P_i = P_{i-1} \cup \{ B \in V \mid \exists\, B \to \alpha \in R \quad \alpha \in (P_{i-1} \cup \Sigma)^* \}$$

until $P_i = P_{i-1}$.

return P_i.

e.g. $A \to ab$

$B \to aAb$

$C \to aABA$

Example: G: $S \to AC \mid BS \mid B$

$A \to aA \mid aF$

$B \to CF \mid b$

$C \to aC \mid Da$

$D \to aD \mid BD \mid C$

$E \to aA \mid BSA$

$F \to bB \mid b$

$P_1 = \{B, F\}$

$P_2 = \{B, F, S, A\}$

$P_3 = \{B, F, S, A, E\}$

$P_4 = \{B, F, S, A, E\} = P_3$

$\Rightarrow$ C, D are not productive.

Q: How to eliminate C and D?

Remove all the rules that involve them.

G': $S \to BS \mid B$

$A \to aA \mid aF$

$B \to b$

$E \to aA \mid BSA$

$F \to bB \mid b$

Note:

$L(G) = L(G')$

Elimination of unreachable symbols:

Idea: Find all the reachable symbols first.

Algorithm:

Input: A CFG $G = (V, \Sigma, R, S)$

Output: $U \subseteq V \cup \Sigma$, the set of reachable symbols

- $U_0 = \{S\}$
- $i = 0$
- repeat
  - $i = i+1$
  - $U_i = U_{i-1} \cup \{X \in V \cup \Sigma \mid \exists A \to \alpha X \beta \in R, A \in U_{i-1}\}$.
- until $U_i = U_{i-1}$
- Return $U = U_i$.

Example:

G: $S \to AC \mid BS \mid B$

$A \to aA \mid aF$

$B \to CF \mid b$

$C \to aC \mid D$

$D \to aD \mid BD \mid C$

$E \to aA \mid BSA$

$F \to bB \mid b$

G': $S \to BS \mid B$

$A \to aA \mid aF$

$B \to b$

$E \to aA \mid BSA$

$F \to bB \mid b$

$P_1 = \{B, F\}$

$P_2 = \{B, F, S, A\}$

$P_3 = \{B, F, S, A, E\}$

$P_4 = \{B, F, S, A, E\} = P_3$

$\Rightarrow$ C, D are not productive.

G': $S \to BS \mid B$

$A \to aA \mid aF$

$B \to b$

$E \to aA \mid BSA$

$F \to bB \mid b$

Let us find reachable symbols in G':

$U_0 = \{S\}$

$U_1 = \{S, B\}$

$U_2 = \{S, B, b\}$

$U_3 = \{S, B, b\}$

$\Rightarrow$ A, E, F, a are not reachable.

Now remove all unreachable symbols from G' by deleting the rules that involve them.

G'': $S \to BS \mid B$

$B \to b$

Note:

$L(G) = L(G') = L(G'') = L(bb^*)$

Example: G: $S \to ABC \mid Bab$

$A \to aA \mid BaC \mid aaa$

$B \to bBb \mid a$

$C \to CA \mid AC$

Prod. nonterminals:

$P_1 = \{A, B\}$

$P_2 = \{A, B, S\}$

$P_3 = P_2$

$\Rightarrow$ C is not productive.

Remove C:

G': $S \to Bab$

$A \to aA \mid aaa$

$B \to bBb \mid a$

Reachable Symbols in G':

$V_0 = \{S\}$

$V_1 = \{S, B, a, b\}$

$V_2 = V_1$

$\Rightarrow$ A is unreachable.

Remove A from G'

G'': $S \to Bab$

$B \to bBb \mid a$

Remark: From a given CFG G, we construct a CFG G' with only useful symbols s.t. $L(G) = L(G')$ by

1) Removing all nonproductive nonterminals
2) Removing all unreachable symbols.

Note that applying step 2 first may lead to incorrect results.

Example: G: $S \to AB \mid a$

$A \to a$

- All symbol are reachable: No change.
- B is not productive:

$\Rightarrow$ G': $S \to a$, $A \to a$ but still we have A, which is unreachable!

Exercise:

G: $S \to AB \mid AC$

$A \to aAb \mid bAa \mid a$

$B \to bbA \mid aaB \mid AB$

$C \to abCa \mid aDb$

$D \to bD \mid aC$

$E \to bE \mid \varepsilon$