Lecture 18

CHOMSKY NORMAL FORM (CNF)

A CFG G is in CNF if every rule in G has the form:

- $A \to BC$ or $(A, B, C \in V)$
- $A \to a$ $(a \in \Sigma, A \in V)$

(Assume: All symbols are useful and $\varepsilon \notin L(G)$)

Theorem: Given a CFG G, there is an algorithm to convert it into its CNF.

i.e. $G \rightsquigarrow G'$ s.t. $L(G) = L(G')$, G' in CNF

Proof: W.l.o.g. we assume that G has no ε-productions, no chain rules and no useless symbols. Then we apply two more steps:

Step 1: For each $a \in \Sigma$, define a new nonterminal D_a and a new rule $D_a \to a$. Then, replace each rule $A \to \alpha$, $|\alpha| \geq 2$,

with a new rule $A \to \bar{\alpha}$ s.t. $\bar{\alpha}$ is obtained from α by replacing each $a \in \Sigma$ in it with D_a.

Step 2: Let $A \to A_1 \cdot A_2 \cdot \dots A_m$, $m \geq 3$, be a rule. Replace it by $A \to A_1 C_1$, $C_1 \to A_2 C_2$, ... , $C_{m-2} \to A_{m-1} A_m$.

where $C_1, \dots, C_{m-2}$ are new nonterminals. ∎

e.g. $A \to BCD \rightsquigarrow A \to BE_1$, $E_1 \to CD$.

Example: G: $S \to AACD$
$A \to aA \mid \varepsilon$
$C \to aC \mid a$
$D \to aD \mid Db \mid \varepsilon$

① **Remove ε-productions:**

A, D are nullable.

G': $S \to AACD \mid ACD \mid AAC \mid CD \mid AC \mid C$
$A \to aA \mid a$
$C \to aC \mid a$
$D \to aD \mid a \mid Db \mid b$

② Remove chain rules:

$CHAIN(S) = \{C\}$

G'': $S \to AACD \mid ACD \mid AAC \mid CD \mid AC \mid aC \mid a$

$A \to aA \mid a$

$C \to aC \mid a$

$D \to aD \mid a \mid Db \mid b$

③ Remove useless symbols: No change.

- All nonterminals are productive.
- All symbols are reachable

④ Convert to CNF:

Step 1:

G''': $S \to AACD \mid ACD \mid AAC \mid CD \mid AC \mid D_a C \mid a$

$A \to D_a A \mid a$

$C \to D_a C \mid a$

$D \to D_a D \mid a \mid D D_b \mid b$

$D_a \to a$

$D_b \to b$

Step 2:

$G^{(iv)}$: $S \to CD \mid AC \mid D_a C \mid a$

$S \to AE_1 \quad E_1 \to AE_2 \quad E_2 \to CD$

$S \to AF_1 \quad F_1 \to CD$

$S \to A G_1 \quad G_1 \to AC$

$A \to D_a A \mid a$

$C \to D_a C \mid a$

$D \to D_a D \mid a \mid D D_b \mid b$

$D_a \to a$

$D_b \to b$

Example: G: $S \to ACA$

$A \to aAa \mid B \mid C$

$B \to bB \mid b$

$C \to cC \mid \varepsilon$

① Remove ε-productions:

C, A, S are nullable.

$N_0 = \{C\}$

$N_1 = \{C, A\}$

$N_2 = \{C, A, S\}$

$N_3 = N_2$

G': $S \to ACA \mid CA \mid AA \mid AC \mid A \mid C$

$A \to aAa \mid aa \mid B \mid C$

$B \to bB \mid b$

$C \to cC \mid c$

② Remove chain rules:

$CHAIN(S) = \{C, A, B\}$

$CHAIN(A) = \{B, C\}$

G'': $S \to ACA \mid CA \mid AA \mid AC \mid aAa \mid aa \mid cC \mid c \mid bB \mid b$

$A \to aAa \mid aa \mid cC \mid c \mid bB \mid b$

$B \to bB \mid b$

$C \to aC \mid c$

③ Remove useless symbols: No change. (Why?)

④ Convert to CNF:

Step 1:

G''': $S \to ACA \mid CA \mid AA \mid AC \mid D_aAD_a \mid D_aD_a \mid D_cC \mid c \mid D_bB \mid b$

$A \to D_aAD_a \mid D_aD_a \mid D_cC \mid c \mid D_bB \mid b$

$B \to D_bB \mid b$

$C \to D_cC \mid c$

$D_a \to a$

$D_b \to b$

$D_c \to c$

Step 2:

G''': $S \to CA \mid AA \mid AC \mid D_aD_a \mid D_cC \mid c \mid D_bB \mid b$

$S \to AE_1 \quad E_1 \to CA$

$S \to D_aF_1 \quad F_1 \to AD_a$

$A \to D_aD_a \mid D_cC \mid c \mid D_bB \mid b$

$A \to D_a G_1 \quad G_1 \to A D_a$

$B \to D_b B \mid b$

$C \to D_c C \mid c$

$D_a \to a$

$D_b \to b$

$D_c \to c$

Example: G: $S \to ABCBCD$

$A \to CD$

$B \to Cb$

$C \to A \mid \varepsilon$

$D \to bD \mid \varepsilon$

$N_0 = \{C, D\}$

$N_1 = \{C, D, A\}$

$N_2 = N_1$

$\Rightarrow$ C, D, A are nullable.

① Remove ε-productions:

G': $S \to ABCBCD \mid BCBCD \mid ABBCD \mid ABCBD \mid ABCBC \mid$

$BBCD \mid BCBD \mid BCBC \mid ABBD \mid ABBC \mid ABCB$

$BBD \mid BBC \mid BCD \mid ABB \mid BB$

$A \to CD \mid C \mid D$

$B \to Cb \mid b$

$C \to A$

$D \to bD \mid b$

② Remove chain rules:

CHAIN(A) = {C, D}

CHAIN(C) = {A, D}

G'': S → ABCBCD | BCBCD | ABBCD | ABCBD | ABCBC |
BBCD | BCBD | BCBC | ABBD | ABBC | ABCB
BBD | BBC | BCD | ABB | BB

A → CD | bD | b

B → Cb | b

C → CD | bD | b

D → bD | b

③ Remove useless symbols: No change.

④ Convert to CNF

Step 1:

G''': S → ABCBCD | BCBCD | ABBCD | ABCBD | ABCBC |
BBCD | BCBD | BCBC | ABBD | ABBC | ABCB
BBD | BBC | BCD | ABB | BB

A → CD | D_bD | b

B → Cb | b

$C \to CD \mid D_b D \mid b$

$D \to D_b D \mid b$

$D_b \to b$

Step 2:

$G^{(iv)}$: $S \to AA_1$ $A_1 \to BA_2$ $A_2 \to CA_3$ $A_2 \to BA_4$ $A_4 \to CD$

$S \to BB_1$ $B_1 \to CB_2$ $B_2 \to BB_3$ $B_3 \to CD$

$S \to AC_1$ $C_1 \to BC_2$ $C_2 \to BC_3$ $C_3 \to CD$

$S \to AD_1$ $D_1 \to BD_2$ $D_2 \to CD_3$ $D_3 \to BD$

$S \to AE_1$ $E_1 \to BE_2$ $E_2 \to CE_3$ $E_3 \to BC$

$S \to BF_1$ $F_1 \to BF_2$ $F_2 \to CD$

$S \to BG_1$ $G_1 \to CG_2$ $G_2 \to BD$

$S \to BH_1$ $H_1 \to CH_2$ $H_2 \to BC$

$S \to AJ_1$ $J_1 \to BJ_2$ $J_2 \to BD$

$S \to AK_1$ $K_1 \to BK_2$ $K_2 \to BC$

$S \to AL_1$ $L_1 \to BL_2$ $L_2 \to CB$

$S \to BM_1$ $M_1 \to BD$

$S \to BN_1$ $N_1 \to BC$

$S \to BP_1$ $P_1 \to CB$

$S \rightarrow AQ_1 \qquad Q_1 \rightarrow BB$

$S \rightarrow BB$

$A \rightarrow CD \mid D_bD \mid b$

$B \rightarrow Cb \mid b$

$C \rightarrow CD \mid D_bD \mid b$

$D \rightarrow D_bD \mid b$

$D_b \rightarrow b$