Quiz

unwanted in CFG:
– variables not used in successful derivations $S \Rightarrow^* x \in \Sigma^*$

CFG $G = (V, \Sigma, S, P)$

Definition

variable A is *live* if $A \Rightarrow^* x$ for some $x \in \Sigma^*$.

variable A is *reachable* if $S \Rightarrow^* \alpha A \beta$ for some $\alpha, \beta \in (\Sigma \cup V)^*$.

variable A is *useful* if there is a derivation of the form $S \Rightarrow^* \alpha A \beta \Rightarrow^* x$ for some string $x \in \Sigma^*$.

useful implies live and reachable.

conversely, . . .

[M] Exercise 4.51, 4.52, 4.53

CFG $G = (V, \Sigma, S, P)$

Definition

variable A is *live* if $A \Rightarrow^* x$ for some $x \in \Sigma^*$.

variable A is *reachable* if $S \Rightarrow^* \alpha A \beta$ for some $\alpha, \beta \in (\Sigma \cup V)^*$.

variable A is *useful* if there is a derivation of the form $S \Rightarrow^* \alpha A \beta \Rightarrow^* x$ for some string $x \in \Sigma^*$.

useful implies live and reachable.
For $S \to AB \mid b$ and $A \to a$, variable A is live and reachable, not useful.
[M] Exercise 4.51, 4.52, 4.53

Live variables

Construction

- $N_0 = \varnothing$
- $N_{i+1} = N_i \cup \{ A \in V \mid A \to \alpha \text{ in } P, \text{ with } \alpha \in (N_i \cup \Sigma)^* \}$

$N_1 = \{ A \in V \mid A \to x \text{ in } P, \text{ with } x \in \Sigma^* \}$
$N_0 \subseteq N_1 \subseteq N_2 \subseteq \cdots \subseteq V$
there exists a k such that $N_k = N_{k+1}$
A is live iff $A \in \bigcup_{i \geq 0} N_i = N_k$
(minimal) depth of derivation tree $A \Rightarrow^* x$

Live variables

Construction

- $N_0 = \varnothing$
- $N_{i+1} = N_i \cup \{\ A \in V \mid A \to \alpha \text{ in } P, \text{ with } \alpha \in (N_i \cup \Sigma)^* \ \}$

Exercise 4.53(c_i).

$S \to ABC \mid BaB \qquad A \to aA \mid BaC \mid aaa$

$B \to bBb \mid a \qquad C \to CA \mid AC$

Reachable variables

Construction

- $N_0 = \{S\}$
- $N_{i+1} = N_i \cup \{\ A \in V \mid B \to \alpha_1 A \alpha_2 \text{ in } P,\ \text{with } B \in N_i\ \}$

$N_0 \subseteq N_1 \subseteq N_2 \subseteq \cdots \subseteq V$
there exists a k such that $N_k = N_{k+1}$
A is reachable iff $A \in \bigcup_{i \geq 0} N_i = N_k$
(minimal) length of derivation $S \Rightarrow^* \alpha A \beta$

Reachable variables

Construction

- $N_0 = \{S\}$
- $N_{i+1} = N_i \cup \{\ A \in V \mid B \to \alpha_1 A \alpha_2 \text{ in } P, \text{ with } B \in N_i\ \}$

$N_0 \subseteq N_1 \subseteq N_2 \subseteq \cdots \subseteq V$
there exists a k such that $N_k = N_{k+1}$
A is reachable iff $A \in \bigcup_{i \geq 0} N_i = N_k$
(minimal) length of derivation $S \Rightarrow^* \alpha A \beta$

- remove all non-live variables (and productions that contain them)
- remove all unreachable variables (and their productions)

then all variables are useful

Reachable variables

Construction

- $N_0 = \{S\}$
- $N_{i+1} = N_i \cup \{\ A \in V \mid B \rightarrow \alpha_1 A \alpha_2 \text{ in } P, \text{ with } B \in N_i\ \}$

Exercise 4.53(c_i)., ctd

$S \rightarrow BaB \qquad A \rightarrow aA \mid aaa \qquad B \rightarrow bBb \mid a$

– remove all non-live variables (and productions that contain them)
– remove all unreachable variables (and productions)

then all variables are useful

does not work the other way around ...

Exercise 4.53(c_i)., revisited

$S \to ABC \mid BaB \qquad A \to aA \mid BaC \mid aaa$

$B \to bBb \mid a \qquad C \to CA \mid AC$

unwanted in CFG:
– variables not used in successful derivations $S \Rightarrow^* x \in \Sigma^*$

sometimes unwanted:
– $A \to B$ A, B variables unit productions [chain rules]

$$S \to A \mid aB \qquad A \to B \mid bS \qquad B \to bb \mid S$$

$$S \Rightarrow A \Rightarrow B \Rightarrow S$$

unwanted in CFG:
– variables not used in successful derivations $\quad S \Rightarrow^* x \in \Sigma^*$
– $A \to B \quad A, B$ variables $\quad$ unit productions [chain rules]

sometimes unwanted:
– $A \to \Lambda \quad A$ variable $\quad \Lambda$-productions

$$S \to AB \mid aB \qquad A \to BS \mid bS \qquad B \to bb \mid \Lambda$$

$$S \Rightarrow AB \Rightarrow BSB \Rightarrow SB \Rightarrow S$$

Let l be length of a string in a derivation
Let t be number of terminals in a string in a derivation

If G has no Λ-productions, and no unit productions,
then . . .

Let l be length of a string in a derivation
Let t be number of terminals in a string in a derivation

If G has no Λ-productions, and no unit productions,
then $l + t$ strictly increases in every step of a derivation
Proof ...

Hence, a string $x \in \Sigma^*$ can only be generated in derivations of at most $2|x| - 1$ steps
May be used to test if $x \in L(G)$

unwanted in CFG:
– variables not used in successful derivations $S \Rightarrow^* x \in \Sigma^*$
– $A \to \Lambda$ A variable Λ-productions
– $A \to B$ A, B variables unit productions [chain rules]

restricted CFG, with 'nice' form
Chomsky normal form $A \to BC$, $A \to \sigma$
Greibach normal form (⊠) $A \to \sigma B_1 \dots B_k$

Idea:

Example

$A \to BCDCB$
$B \to b \mid \Lambda$
$C \to c \mid \Lambda$
$D \to d$

Definition

variable A is *nullable* iff $A \Rightarrow^* \Lambda$

Theorem

– *if $A \to \Lambda$ then A is nullable*

– *if $A \to B_1 B_2 \dots B_k$ and all B_i are nullable, then A is nullable*

[M] Def 4.26 / Exercise 4.48

Construction

– $N_0 = \varnothing$

– $N_{i+1} = N_i \cup \{\ A \in V \mid A \to \alpha \text{ in } P, \text{ with } \alpha \in N_i^*\ \}$

$N_1 = \{\ A \in V \mid A \to \Lambda \text{ in } P\ \}$

$N_0 \subseteq N_1 \subseteq N_2 \subseteq \cdots \subseteq V$

there exists a k such that $N_k = N_{k+1}$

A is nullable iff $A \in \bigcup_{i \geq 0} N_i = N_k$

Construction

– identify nullable variables
– for every production $A \to \alpha$ add $A \to \beta$,
  where β is obtained from α by removing one or more nullable variables
– remove all Λ-productions

Grammar for $\{ a^i b^j c^k \mid i = j \text{ or } i = k \}$

$S \to TU \mid V$
$T \to aTb \mid \Lambda$
$U \to cU \mid \Lambda$
$V \to aVc \mid W$
$W \to bW \mid \Lambda$

Grammar for $\{ a^i b^j c^k \mid i = j \text{ or } i = k \}$

$S \to TU \mid V$
$T \to aTb \mid \Lambda$
$U \to cU \mid \Lambda$
$V \to aVc \mid W$
$W \to bW \mid \Lambda$

$N_1 = \{T, U, W\}$, variables with Λ at right-hand side productions

$N_2 = \{T, U, W\} \cup \{S, V\}$, variables with $\{T, U, W\}^*$ at rhs productions

$N_3 = N_2 = \{T, U, W, S, V\}$, all variables found, no new add all productions, where (any number of) nullable variables are removed...

$$
\begin{aligned}
S &\to TU \mid V \\
T &\to aTb \mid \Lambda \\
U &\to cU \mid \Lambda \\
V &\to aVc \mid W \\
W &\to bW \mid \Lambda
\end{aligned}
$$

[M] Ex. 4.31

add all productions, where (any number of) nullable variables are removed

$S \to TU \mid V$	$S \to T \mid U \mid \Lambda$
$T \to aTb \mid \Lambda$	$T \to ab$
$U \to cU \mid \Lambda$	$U \to c$
$V \to aVc \mid W$	$V \to ac \mid \Lambda$
$W \to bW \mid \Lambda$	$W \to b$

remove all Λ-productions...

[M] Ex. 4.31

add all productions, where (any number of) nullable variables are removed

$$\begin{array}{ll} S \to TU \mid V & S \to T \mid U \mid \Lambda \\ T \to aTb \mid \Lambda & T \to ab \\ U \to cU \mid \Lambda & U \to c \\ V \to aVc \mid W & V \to ac \mid \Lambda \\ W \to bW \mid \Lambda & W \to b \end{array}$$

remove all Λ-productions

$$\begin{array}{l} S \to TU \mid V \mid T \mid U \\ T \to aTb \mid ab \\ U \to cU \mid c \\ V \to aVc \mid W \mid ac \\ W \to bW \mid b \end{array}$$

[M] Ex. 4.31

Can we still derive $\{ a^i b^j c^k \mid i = j \text{ or } i = k \}$?

Theorem

For every CFG G *there is* CFG G_1 *without* Λ*-productions such that* $L(G_1) = L(G) - \{\Lambda\}$.

Proof $L(G) - \{\Lambda\} \subseteq L(G_1)$ ...

[M] Thm 4.27

Theorem

For every CFG G *there is* CFG G_1 *without* Λ*-productions such that* $L(G_1) = L(G) - \{\Lambda\}$.

Proof $L(G) - \{\Lambda\} \subseteq L(G_1)$

$G = (V, \Sigma, S, P)$
Consider arbitrary $x \in L(G) - \{\Lambda\}$
$S \Rightarrow_G^* x$, i.e., $S \Rightarrow_G^n x$ for some $n \geq 1$
Needed: $S \Rightarrow_{G_1}^* x$
We prove more general statement:
For all $A \in V$, $n \geq 1$ and $x \in \Sigma^* - \{\Lambda\}$, if $A \Rightarrow_G^n x$, then $A \Rightarrow_{G_1}^* x$, using induction on n
Basis, $n = 1$: If $A \Rightarrow_G x$, then also $A \Rightarrow_{G_1} x$

Theorem

For every CFG G *there is* CFG G_1 *without* Λ*-productions such that* $L(G_1) = L(G) - \{\Lambda\}$.

Proof $L(G) - \{\Lambda\} \subseteq L(G_1)$ (continued)

Induction hypothesis: Let $k \geq 1$, and suppose that for all $A \in V$, $n \leq k$ and $x \in \Sigma^* - \{\Lambda\}$, if $A \Rightarrow_G^n x$, then $A \Rightarrow_{G_1}^* x$
Induction step: Consider $A \Rightarrow_G^{k+1} x$
then $A \Rightarrow_G X_1 X_2 \dots X_m \Rightarrow_G^k x = x_1 x_2 \dots x_m$, for some $m \geq 1$ and $X_1, X_2, \dots, X_m \in V \cup \Sigma$
Three cases:
1. X_i is terminal
2. X_i is variable and $x_i \neq \Lambda$
3. X_i is variable and $x_i = \Lambda$

[M] Thm 4.27

Idea:

Example

$A \rightarrow B \mid aCb$
$B \rightarrow C \mid Bb \mid Bc$
$C \rightarrow c \mid ABC$

Assume Λ-productions have been removed

Variable B is *A-derivable*, if
- $B \neq A$, and
- $A \Rightarrow^* B$ (using only unit productions)

Construction

- $N_1 = \{ B \in V \mid B \neq A \text{ and } A \to B \text{ in } P \}$
- $N_{i+1} = N_i \cup \{ C \in V \mid C \neq A \text{ and } B \to C \text{ in } P, \text{ with } B \in N_i \}$

$N_1 \subseteq N_2 \subseteq \cdots \subseteq V$
there exists a k such that $N_k = N_{k+1}$
B is A-derivable iff $B \in \bigcup_{i \geq 1} N_i = N_k$

Construction

- for each $A \in V$, identify A-derivable variables
- for every pair (A, B) where B is A-derivable,
  and every production $B \to \alpha$ add $A \to \alpha$
- remove all unit productions

Grammar for $\{ a^i b^j c^k \mid i = j \text{ or } i = k \}$

$$
\begin{aligned}
S &\to TU \mid V \mid T \mid U \\
T &\to aTb \mid ab \\
U &\to cU \mid c \\
V &\to aVc \mid W \mid ac \\
W &\to bW \mid b
\end{aligned}
$$

$$
\begin{aligned}
S &\to TU \mid V \mid T \mid U \\
T &\to aTb \mid ab \\
U &\to cU \mid c \\
V &\to aVc \mid W \mid ac \\
W &\to bW \mid b
\end{aligned}
$$

S-derivable: $\{V, T, U\}, \{V, T, U, W\}$ V-derivable: $\{W\}$

New productions:

$S \to aTb \mid ab \quad S \to cU \mid c \quad S \to aVc \mid W \mid ac \quad S \to bW \mid b$

$V \to bW \mid b$

Remove unit productions:

$$
\begin{aligned}
S &\to TU \mid aTb \mid ab \mid cU \mid c \mid aVc \mid ac \mid bW \mid b \\
T &\to aTb \mid ab \\
U &\to cU \mid c \\
V &\to aVc \mid ac \mid bW \mid b \\
W &\to bW \mid b
\end{aligned}
$$

Definition

CFG in *Chomsky normal form*

productions are of the form

– $A \to BC$ variables A, B, C

– $A \to \sigma$ variable A, terminal σ

Theorem

For every CFG G *there is* CFG G_1 *in CNF such that* $L(G_1) = L(G) - \{\Lambda\}$.

[M] Def 4.29, Thm 4.30

Construction

1. remove Λ-productions
2. remove unit productions
3. introduce variables for terminals $\quad X_\sigma \to \sigma$
4. split long productions

$$A \to aBabA$$

is replaced by

$$X_a \to a \qquad X_b \to b \qquad A \to X_a B X_a X_b A$$

$$A \to ACBA$$

is replaced by

$$A \to AY_1 \qquad Y_1 \to CY_2 \qquad Y_2 \to BA$$

Mind the order

Grammar for $\{ a^i b^j c^k \mid i = j \text{ or } i = k \}$

$S \to TU \mid V$

$T \to aTb \mid \Lambda \qquad U \to cU \mid \Lambda$

$V \to aVc \mid W \qquad W \to bW \mid \Lambda$

After removing Λ-productions and unit productions, we obtain (see before)

$S \to TU \mid aTb \mid ab \mid cU \mid c \mid aVc \mid ac \mid bW \mid b$

$T \to aTb \mid ab \qquad U \to cU \mid c$

$V \to aVc \mid ac \mid bW \mid b \qquad W \to bW \mid b$

Now introduce productions for the terminals. . .

Grammar for $\{ a^i b^j c^k \mid i = j \text{ or } i = k \}$

$S \to TU \mid V$

$T \to aTb \mid \Lambda \qquad U \to cU \mid \Lambda$

$V \to aVc \mid W \qquad W \to bW \mid \Lambda$

After removing Λ-productions and unit productions, we obtain (see before)

$S \to TU \mid aTb \mid ab \mid cU \mid c \mid aVc \mid ac \mid bW \mid b$

$T \to aTb \mid ab \qquad U \to cU \mid c$

$V \to aVc \mid ac \mid bW \mid b \qquad W \to bW \mid b$

Now introduce productions for the terminals:

$X_a \to a \qquad X_b \to b \qquad X_c \to c$

$S \to TU \mid X_a T X_b \mid X_a X_b \mid X_c U \mid c \mid X_a V X_c \mid X_a X_c \mid X_b W \mid b$

$T \to X_a T X_b \mid X_a X_b$

$U \to X_c U \mid c$

$V \to X_a V X_c \mid X_a X_c \mid X_b W \mid b$

$W \to X_b W \mid b$

Only a few productions that are too long:

$S \to X_a T X_b \mid X_a V X_c$
$T \to X_a T X_b$
$V \to X_a V X_c$

Split these long productions...

Only a few productions that are too long:

$S \to X_a T X_b \mid X_a V X_c$
$T \to X_a T X_b$
$V \to X_a V X_c$

Split these long productions:

$S \to X_a Y_1 \mid X_a Y_2$
$Y_1 \to T X_b \qquad Y_2 \to V X_c$
$T \to X_a Y_3$
$V \to X_a Y_4$
$Y_3 \to T X_b \qquad Y_4 \to V X_c$

Note that we can reuse Y_1 ($\approx Y_3$) and Y_2 ($\approx Y_4$) for two productions

Homework 3! (probably Wednesday)