Context-Free Grammars - Chomsky Normal Form

Lecture 16
Section 2.1

Robb T. Koether

Hampden-Sydney College

Wed, Oct 1, 2008

Outline

Homework Review

Exercise 3, page 128.

Answer each part for the following context-free grammar G.

$$
\begin{aligned}
R &\rightarrow XRX \mid S \\
S &\rightarrow \mathbf{a}T\mathbf{b} \mid \mathbf{b}T\mathbf{a} \\
T &\rightarrow XTX \mid X \mid \varepsilon \\
X &\rightarrow \mathbf{a} \mid \mathbf{b}
\end{aligned}
$$

(a) What are the variables of G?

(b) What are the terminals of G?

(c) Which is the start variable of G?

Homework Review

Solution

(a) The variables are $\{R, S, T, X\}$.

(b) The terminals are $\{\mathbf{a}, \mathbf{b}\}$.

(c) The start symbol is R.

Homework Review

Exercise 3, page 128.

(d) Give three strings in $L(G)$.

(e) Give three strings *not* in $L(G)$.

(f) True or False: $T \Rightarrow \mathbf{aba}$.

(g) True or False: $T \overset{*}{\Rightarrow} \mathbf{aba}$.

(h) True or False: $T \Rightarrow T$.

(i) True or False: $T \overset{*}{\Rightarrow} T$.

Homework Review

Solution

(d) $\mathbf{ab}, \mathbf{ba}, \mathbf{aab} \in L(G)$.

(e) $\mathbf{a}, \mathbf{b}, \varepsilon \notin L(G)$.

(f) False, $T \not\Rightarrow \mathbf{aba}$.

(g) True, $T \overset{*}{\Rightarrow} \mathbf{aba}$.

(h) False, $T \not\Rightarrow T$.

(i) True, $T \overset{*}{\Rightarrow} T$.

Homework Review

Exercise 3, page 128.

(j) True or False: $XXX \stackrel{*}{\Rightarrow} \mathbf{aba}$.

(k) True or False: $X \stackrel{*}{\Rightarrow} \mathbf{aba}$.

(l) True or False: $T \stackrel{*}{\Rightarrow} XX$.

(m) True or False: $T \stackrel{*}{\Rightarrow} XXX$.

(n) True or False: $S \stackrel{*}{\Rightarrow} \varepsilon$.

Homework Review

Solution

(j) True, $XXX \stackrel{*}{\Rightarrow} \mathbf{aba}$.

(k) False, $X \not\Rightarrow^* \mathbf{aba}$.

(l) True, $T \stackrel{*}{\Rightarrow} XX$.

(m) True, $T \stackrel{*}{\Rightarrow} XXX$.

(n) False, $S \not\Rightarrow^* \varepsilon$.

Homework Review

Exercise 3, page 128.

(o) Give a description in English of $L(G)$.

Homework Review

Solution

(o) The R must eventually produce X^nSX^n, for some $n \geq 0$.

(p) The S must produce either $\mathbf{a}T\mathbf{b}$ or $\mathbf{b}T\mathbf{a}$.

(q) The T must eventually produce X^m, for some $m \geq 0$.

(r) Thus, so far, R produces $X^n\mathbf{a}X^m\mathbf{b}X^n$ or $X^n\mathbf{b}X^m\mathbf{a}X^n$, for some $m, n \geq 0$.

Homework Review

Solution

(o) Now X can be replaced with either $\mathbf{a}$ or $\mathbf{b}$

(p) Therefore, X^n and X^m can be any string in Σ^*.

(q) Thus, the language is the set of all strings of the form $u\mathbf{a}v\mathbf{b}w$ or $u\mathbf{b}v\mathbf{a}w$, where $u, v, w \in \Sigma^*$ and $|u| = |w|$.

(r) It is not hard to see that this is the complement of the set $\{ww^R \mid w \in \Sigma^*\}$.

Chomsky Normal Form

Definition (Chomsky Normal Form)

A grammar is in Chomsky Normal Form, abbreviated CNF, if each rule is of the form

- $A \to BC$, or
- $A \to a$,

where B and C are nonteriminals not equal to S and a is a terminal. Furthermore, the rule $S \to \varepsilon$ is allowed.

Chomsky Normal Form

Theorem (Chomsky Normal Form)

Every context-free language is generated by a grammar in Chomsky Normal Form.

Chomsky Normal Form

Outline of proof.

Begin with a grammar for the context-free language.

- Add a new start symbol S_0.
- Eliminate all ε-rules $A \to \varepsilon$.
- Eliminate all unit rules $A \to B$.
- Eliminate all mixed rules.
- Eliminate all long rules.

□

New Start Symbol

Proof (New start symbol S_0).

- Add the rule $S_0 \to S$.

□

Example

Example (New start symbol S_0)

- Start with the grammar

$$\begin{aligned} S &\to SXS \mid \varepsilon \\ X &\to \mathbf{ab} \mid \varepsilon \end{aligned}$$

- Add the rule

$$S_0 \to S$$

Example

Example (New start symbol S_0)

- We now have the grammar

$$\begin{array}{rcl} S_0 & \to & S \\ S & \to & SXS \mid \varepsilon \\ X & \to & \mathbf{ab} \mid \varepsilon \end{array}$$

Eliminate All ε-Rules

Proof (Eliminate all ε-rules).

- For each rule $A \to \varepsilon$ and each rule $B \to uAv$ (with A on the right), add a rule $B \to uv$.
- Eliminate the rule $A \to \varepsilon$.

□

Example

Example (Eliminate all ε-rules)

- Apply the rules $S \to \varepsilon$ and $X \to \varepsilon$ to the other rules, creating the rules

$$\begin{aligned} S_0 &\to \varepsilon \\ S &\to X \\ S &\to SS \\ S &\to XS \\ S &\to SX \\ S &\to S \end{aligned}$$

(Don't bother keeping the last rule.)

Example

Example (Eliminate all ε-rules)

- Eliminate the rules

$$
\begin{aligned}
S &\rightarrow \varepsilon \\
X &\rightarrow \varepsilon
\end{aligned}
$$

Example

Example (Eliminate all ε-rules)

- We now have

$$
\begin{aligned}
S_0 &\rightarrow S \mid \varepsilon \\
S &\rightarrow SXS \mid SS \mid SX \mid XS \mid X \\
X &\rightarrow \mathbf{ab}
\end{aligned}
$$

Eliminate All Unit Rules

Proof (Eliminate all unit rules).

- If $A \to B$ and $B \to u$ are rules, then add the rule $A \to u$.
- Eliminate the rule $A \to B$.

□

Example

Example (Eliminate all unit rules)

- Add the rules

$$\begin{aligned} S_0 &\rightarrow SXS \mid SS \mid SX \mid XS \mid X \mid \mathbf{ab} \\ S &\rightarrow \mathbf{ab} \end{aligned}$$

- Eliminate the rules

$$\begin{aligned} S_0 &\rightarrow S \\ S &\rightarrow X \end{aligned}$$

Example

Example (Eliminate all unit rules)

- We now have

$$\begin{aligned} S_0 &\rightarrow SXS \mid SS \mid SX \mid XS \mid \mathbf{ab} \mid \varepsilon \\ S &\rightarrow SXS \mid SS \mid SX \mid XS \mid \mathbf{ab} \\ X &\rightarrow \mathbf{ab} \end{aligned}$$

Eliminate All Mixed Rules

Definition (Mixed rule)

A mixed rule is a rule whose right-hand side has length at least 2 and contains at least one terminal.

Proof (Eliminate all mixed rules).

- Add rules

$$A \longrightarrow a$$

for all terminals a appearing in mixed rules.

- Replace a with A in those rules.

□

Example

Example (Eliminate all mixed rules)

- Add the rules

$$\begin{aligned} A &\rightarrow \mathbf{a} \\ B &\rightarrow \mathbf{b} \end{aligned}$$

- Replace the string $\mathbf{ab}$ with AB.

Example

Example (Eliminate all mixed rules)

- We now have

$$
\begin{aligned}
S_0 &\rightarrow SXS \mid SS \mid SX \mid XS \mid AB \mid \varepsilon \\
S &\rightarrow SXS \mid SS \mid SX \mid XS \mid AB \\
X &\rightarrow AB \\
A &\rightarrow \mathbf{a} \\
B &\rightarrow \mathbf{b}
\end{aligned}
$$

Eliminate All Long Rules

Definition (Mixed rule)

A long rule is a rule whose right-hand side has length at least 3.

Eliminate All Long Rules

Proof (Eliminate all long rules).

- Replace the long rules

$$A \rightarrow B_1 B_2 \ldots B_k, (k \geq 3)$$

with

$$\begin{aligned} A &\rightarrow B_1 C_1 \\ C_1 &\rightarrow B_2 C_2 \\ C_2 &\rightarrow B_3 C_3 \\ &\vdots \\ C_{k-2} &\rightarrow B_{k-2} C_{k-2} \\ C_{k-1} &\rightarrow B_{k-1} B_k \end{aligned}$$

□

Example

Example (Eliminate all long rules)

- Replace

$$
\begin{aligned}
S_0 &\rightarrow SXS \\
S &\rightarrow SXS
\end{aligned}
$$

with

$$
\begin{aligned}
S_0 &\rightarrow SY \\
S &\rightarrow SY \\
Y &\rightarrow XS
\end{aligned}
$$

Example

Example (Eliminate all long rules)

- The final result is

$$
\begin{aligned}
S_0 &\rightarrow SY \mid SS \mid SX \mid XS \mid AB \mid \varepsilon \\
S &\rightarrow SY \mid SS \mid SX \mid XS \mid AB \\
X &\rightarrow AB \\
Y &\rightarrow XS \\
A &\rightarrow \mathbf{a} \\
B &\rightarrow \mathbf{b}
\end{aligned}
$$

which is in Chomsky Normal Form.

A Derivation in CNF

Example (A CNF derivation)

- Use this grammar in CNF to derive the string **ababab.**

$$
\begin{aligned}
S_0 &\Rightarrow SY \\
&\Rightarrow SXS \\
&\Rightarrow ABXS \\
&\Rightarrow ABABS \\
&\Rightarrow ABABAB
\end{aligned}
$$

A Derivation in CNF

Example (A CNF derivation)

$$\begin{aligned}
& \vdots \\
\Rightarrow \quad & \mathbf{a}BABAB \\
\Rightarrow \quad & \mathbf{ab}ABAB \\
\Rightarrow \quad & \mathbf{aba}BAB \\
\Rightarrow \quad & \mathbf{abab}AB \\
\Rightarrow \quad & \mathbf{ababa}B \\
\Rightarrow \quad & \mathbf{ababab}.
\end{aligned}$$

CNF Derivations

Theorem

If a grammar G is in CNF and a string w in $L(G)$ has length n, then w is derived from G in exactly $2n - 1$ steps.

The Membership Problem for CFGs

Definition (The Membership Problem for CFGs)

Given a CFG G and a string w, can w be derived from G?

- The previous theorem allows us to solve the Membership Problem.

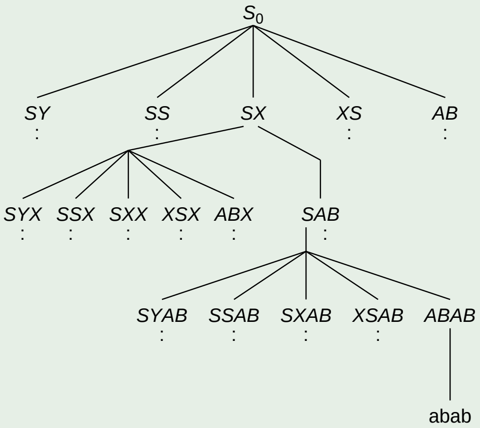
Example

Example (The Membership Problem for CFGs)

- Show that the string **abba** is *not* derivable from the grammar of the previous example.
- Draw a tree of all possible derivations of strings up to length 4.
- This will involve up to 7 steps (but no more).

Example

Example (The Membership Problem for CFGs)



Example

Example (The Membership Problem for CFGs)

- Put the grammar

$$E \rightarrow E + E \mid E * E \mid (E) \mid \mathbf{a} \mid \mathbf{b} \mid \mathbf{c}$$

  into CNF.
- Show that the string $\mathbf{c++}$ is not derivable from it.

Assignment

Homework

- Read Section 2.1, pages 106 - 109.
- Exercise 14, page 129.