CS 4410

Automata, Computability, and Formal Language

Dr. Xuejun Liang

Spring 2019

Chapter 6

Simplification of Context-Free Grammars and Normal Forms

1. Methods for Transforming Grammars
    - A Useful Substitution Rule
    - Removing Useless Productions
    - Removing λ-Productions
    - Removing Unit-Productions
2. Two Important Normal Forms
    - Chomsky Normal Form
    - Greibach Normal Form
3. A Membership Algorithm for Context-Free Grammars*

Learning Objectives

At the conclusion of the chapter, the student will be able to:

- Simplify a context-free grammar by removing useless productions
- Simplify a context-free grammar by removing λ-productions
- Simplify a context-free grammar by removing unit-productions
- Determine whether or not a context-free grammar is in Chomsky normal form
- Transform a context-free grammar into an equivalent grammar in Chomsky normal form
- Determine whether or not a context-free grammar is in Greibach normal form
- Transform a context-free grammar into an equivalent grammar in Greibach normal form

A Useful Substitution Rule

Theorem 6.1: Let G=(V, T, S, P) be a context-free grammar. Suppose that P contains a production of the form

$$A \rightarrow x_1 B x_2.$$

Assume that A and B are different variables and that

$$B \rightarrow y_1 \mid y_2 \mid \ldots \mid y_n$$

is the set of all productions in P which have B as the left side. Let $\widehat{G} = (V, T, S, \widehat{P})$ be the grammar in which $\widehat{P}$ is constructed by deleting

$$A \rightarrow x_1 B x_2$$

from P, and adding to it

$$A \rightarrow x_1 y_1 x_2 \mid x_1 y_2 x_2 \mid \ldots \mid x_1 y_n x_2$$

Then

$$L(\widehat{G}) = L(G)$$

Example 6.1: Consider G=({A, B}, {a, b, c}, A, P) with productions

$$A \rightarrow a \mid aaA \mid abBc,$$

$$B \rightarrow abbA \mid b$$

Substitute the variable B.

Removing Useless Productions

Definition 6.1: Let G=(V, T, S, P) be a context-free grammar. A variable A$\in$V is said to be useful if and only if there is at least one w$\in$L(G) such that

$$S \stackrel{*}{\Rightarrow} xAy \stackrel{*}{\Rightarrow} w,$$

with x, y $\in(V\cup T)^*$. In words, a variable is useful if and only if it occurs in at least one derivation. A variable that is not useful is called useless. A production is useless if it involves any useless variable.

Example 6.2, 6.3:
Eliminate useless symbols and productions

S → A,
A → aA | λ,
B → bA.

S → aS | A | C,
A → a,
B → aa,
C → aCb.

Theorem 6.2: Let G=(V, T, S, P) be a context-free grammar. Then there exists an equivalent grammar $\hat{G} = (\hat{V}, \hat{T}, S, \hat{P})$ that does not contain any useless variables and productions.

Removing Useless Productions

Let G=(V, T, S, P) be a context-free grammar. Compute

$$V_1=\{A\in V : A \overset{*}{\Rightarrow} w \in T^*\}.$$

1. Set V_1 to $\emptyset$.
2. Repeat the following step until no more variables are added to V_1.

   For every $A \in V$ for which P has production of the form

   $A \rightarrow x_1x_2\ldots x_n$ with all x_i in $V_1 \cup T$

   Add A to V_1.

Let G=(V, T, S, P) be a context-free grammar. Compute

$$V_2=\{A\in V : S \overset{*}{\Rightarrow} xAy \in (V\cup T)^*\}.$$

1. Set $V_2 = \{S\}$.
2. Repeat the following step until no more variables are added to V_2.

   For every $B \in V$ for which P has production of the form

   $A \rightarrow xBy$ with $x, y \in (T\cup V)^*$ and $A \in V_2$

   Add A to V_2.

Removing λ-Productions

Definition 6.2: Any production of a context-free grammar of the form $A \Rightarrow \lambda$ is called a λ-production. Any variable A for which the derivation $A \overset{*}{\Rightarrow} \lambda$ is possible is called nullable.

Example 6.4:
Eliminate λ-productions
The language $L=\{a^n b^n : n \geq 1\}$

$$S \to aS_1b,\quad S_1 \to aS_1b \mid \lambda \quad \Longrightarrow \quad S \to aS_1b \mid ab,\quad S_1 \to aS_1b \mid ab$$

Theorem 6.3: Let G be a context-free grammar with λ not in L(G). Then there exists an equivalent grammar $\widehat{G}$ having no λ-productions.

Example 6.5:
Eliminate λ-productions

$$\begin{aligned} S &\to ABaC, \\ A &\to BC, \\ B &\to b \mid \lambda, \\ C &\to D \mid \lambda, \\ D &\to d. \end{aligned} \quad \Longrightarrow \quad \begin{aligned} S &\to ABaC \mid BaC \mid AaC \\ &\quad \mid ABa \mid aC \mid Aa \mid Ba \mid a, \\ A &\to B \mid C \mid BC, \\ B &\to b, \\ C &\to D, \\ D &\to d. \end{aligned}$$

Removing λ-Productions

Let G=(V, T, S, P) be a context-free grammar. Compute

$$V_N = \{A \in V : A \overset{*}{\Rightarrow} \lambda\}.$$

1. For all production $A \rightarrow \lambda$, put A in V_N.
2. Repeat the following step until no further variables are added to V_N.

   For all productions

   $$B \rightarrow A_1A_2\ldots A_n,$$

   where $A_1, A_2, \ldots, A_n$ are in V_N, put B into V_N.

Replace all nullable variables with λ in all possible combinations

Example: Assume B and C $\in V_N$. Then

1. $A \rightarrow ByC$ is replaced by $A \rightarrow y \mid yC \mid By$
2. $A \rightarrow BC$ is replaced by $A \rightarrow C \mid B$

Removing Unit-Productions

Definition 6.3: Any production of a context-free grammar of the form A=>B where A, B $\in$ V is called a unit-production.

Theorem 6.4: Let G=(V, T, S, P) be any context-free grammar without λ-productions. Then there exists a context-free grammar $\widehat{G} = (\widehat{V}, \widehat{T}, S, \widehat{P})$ that does not have any unit-productions and that is equivalent to G.

Steps: 1. Put all non-unit productions of P into $\widehat{P}$

2. For A $\overset{*}{=>}$ B, add to $\widehat{P}$

$$A \rightarrow y_1 \mid y_2 \mid \ldots \mid y_n,$$

where B $\rightarrow$ $y_1 \mid y_2 \mid \ldots \mid y_n$ in $\widehat{P}$

Example 6.6: Eliminate unit-productions

$$S \to Aa \mid B, \quad B \to A \mid bb, \quad A \to a \mid bc \mid B$$

$$\Rightarrow$$

$$S \to a \mid bc \mid bb \mid Aa, \quad A \to a \mid bb \mid bc, \quad B \to a \mid bb \mid bc,$$

Removing λ-Productions
Removing Unit-Productions
Removing Useless Productions

Theorem 6.5: Let L be a context-free language that does not contain λ. Then there exists a context-free grammar that generate L and that does not have any useless productions, λ-productions, or unit-productions.

Steps: 1. Remove λ-productions
2. Remove unit-productions
3. Remove useless productions

Chomsky Normal Form

Definition 6.4: A context-free grammar is in Chomsky normal form if all productions are of form

A → BC, or

A → a

where A, B, C are in V, and a is in T.

Example 6.7: In Chomsky normal form

$$S \rightarrow AS,$$
$$A \rightarrow SA \mid b$$

Not in Chomsky normal form

$$S \rightarrow AS \mid AAS,$$
$$A \rightarrow SA \mid aa$$

Theorem 6.6: Any context-free grammar G=(V, T, S, P) with λ ∉ L(G) has an equivalent grammar $\hat{G} = (\hat{V}, \hat{T}, S, \hat{P})$ in Chomsky normal form.

Example 6.8:
Convert the grammar to Chomsky normal form

$$S \rightarrow ABa,$$
$$A \rightarrow aab$$
$$B \rightarrow Ac$$

Greibach Normal Form

Definition 6.5: A context-free grammar is in Greibach normal form if all productions are of form

$$A \to ax$$

where a is in T and x is in V*.

Example 6.9: In Greibach normal form

$$S \to aAB \mid bBB \mid bB,$$
$$A \to aA \mid bB \mid b,$$
$$B \to b$$

Not in Greibach normal form

$$S \to AB,$$
$$A \to aA \mid bB \mid b,$$
$$B \to b$$

Example 6.10: Convert the grammar $S \to abSb \mid aa$ into Greibach normal form

Theorem 6.7: Any context-free grammar G=(V, T, S, P) with $\lambda \notin L(G)$ has an equivalent grammar $\hat{G} = (\hat{V}, \hat{T}, S, \hat{P})$ in Greibach normal form.