Transforming Grammars

Lecture 17
Section 6.1

Robb T. Koether

Hampden-Sydney College

Mon, Oct 3, 2016

Outline

1. The Membership Problem for CFGs
2. Chomsky Normal Form
3. The Algorithm (Part 1)
    - Eliminate all λ-Productions
    - Eliminate all unit productions
    - Eliminate all useless productions
4. Assignment

Outline

The Membership Problem for CFGs

Definition

The Membership Problem for CFGs Given a context-free grammar G and a string w, is $w \in L(G)$? That is, can w be derived from G?

The Membership Problem for CFGs

Example (Membership Problem for CFGs)

- Consider the grammar

$$S \to SAS \mid \mathbf{b}A\mathbf{a}$$
$$A \to \mathbf{a}S \mid S\mathbf{b} \mid \mathbf{ab} \mid S \mid \lambda$$

The Membership Problem for CFGs

Example (Membership Problem for CFGs)

- Consider the grammar

$$S \to SAS \mid \mathbf{b}A\mathbf{a}$$
$$A \to \mathbf{a}S \mid S\mathbf{b} \mid \mathbf{ab} \mid S \mid \lambda$$

- Is **babbba** $\in L(G)$?

The Membership Problem for CFGs

Example (Membership Problem for CFGs)

- Consider the grammar

$$S \to SAS \mid \mathbf{b}A\mathbf{a}$$
$$A \to \mathbf{a}S \mid S\mathbf{b} \mid \mathbf{ab} \mid S \mid \lambda$$

- Is **babbba** $\in L(G)$?
- If it is, then we can prove that by deriving it.

The Membership Problem for CFGs

Example (Membership Problem for CFGs)

- Consider the grammar

$$S \rightarrow SAS \mid \mathbf{b}A\mathbf{a}$$
$$A \rightarrow \mathbf{a}S \mid S\mathbf{b} \mid \mathbf{ab} \mid S \mid \lambda$$

- Is **babbba** $\in L(G)$?
- If it is, then we can prove that by deriving it.
- But if it isn't, then how do we prove that?

Outline

Chomsky Normal Form

Definition (Chomsky Normal Form)

A grammar G 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 nonterminals and a is a terminal. Furthermore, if $\lambda \in L(G)$, then add the rule $S' \to S \mid \lambda$.

The Membership Problem for CFGs

- What is the benefit of having a grammar in Chomsky Normal Form?
- CNF allows us to solve the membership problem for CFGs.

The Membership Problem for CFGs

Example (Membership Problem for CFGs)

- The grammar of the last example, in CNF:

$$S \rightarrow SD \mid CE \mid SS \mid CB$$
$$A \rightarrow BC \mid SC \mid BS$$
$$B \rightarrow \mathbf{a}$$
$$C \rightarrow \mathbf{b}$$
$$D \rightarrow AS$$
$$E \rightarrow AB$$

The Membership Problem for CFGs

Example (Membership Problem for CFGs)

- The grammar of the last example, in CNF:

$$
\begin{aligned}
S &\rightarrow SD \mid CE \mid SS \mid CB \\
A &\rightarrow BC \mid SC \mid BS \\
B &\rightarrow \mathbf{a} \\
C &\rightarrow \mathbf{b} \\
D &\rightarrow AS \\
E &\rightarrow AB
\end{aligned}
$$

- Is **babbba** $\in L(G)$?

Outline

Chomsky Normal Form

Theorem (Chomsky Normal Form)

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

Proof.

Proof by algorithm. □

Chomsky Normal Form

Outline of proof.

Begin with a grammar for the context-free language.

- Eliminate all λ-productions $A \to \lambda$.
- Eliminate all unit productions $A \to B$.
- Eliminate all useless productions, e.g., $A \to AB$ only.
- Eliminate all mixed productions.
- Eliminate all long productions.

□

Outline

Nullable Variables

Definition

A variable A is nullable if it derives λ. That is, if $A \stackrel{*}{\Rightarrow} \lambda$.

Nullable Variables

- To determine all nullable variables,
  - Let $N = \varnothing$.
  - Add to N all variables A for which there is a production $A \to \lambda$.
  - Repeatedly add to N all variables A for which there is a production

    $$A \to B_1 B_2 \cdots B_k$$

    where all $B_i \in N$ until no more such variables can be added.
  - N is the set of nullable variables.

Eliminate All λ-Productions

Proof (Eliminate all λ-productions).

- For each nullable variable A and for each production $B \to uAv$ (with A on the right), add the production $B \to uv$.
- Eliminate all λ-productions.
- (If S is nullable, then add a new start symbol S' and the rule $S' \to S \mid \lambda$.)

□

Eliminate All λ-Productions

Example (Eliminate all λ-productions)

- Eliminate all λ-productions from the following grammar.

$$
\begin{aligned}
S &\to \mathbf{a}S\mathbf{b} \mid SA \mid \mathbf{b}B \\
A &\to \mathbf{a}A \mid SB \mid \lambda \\
B &\to ABA \mid \lambda
\end{aligned}
$$

Outline

Eliminate All Unit Productions

Proof (Eliminate all unit productions).

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

□

Eliminate All Unit Productions

- A complication occurs if, for variables A and B, we have $A \overset{*}{\Rightarrow} B$ and $B \overset{*}{\Rightarrow} A$.
- To address this,
  - Let P' be the set of all non-unit productions in G.
  - Draw a dependency graph using only unit productions.
  - For every A and B for which $A \overset{*}{\Rightarrow} B$ in G, and for every production $B \to w$ in P', add the production $A \to w$ to P'.

Example

Example (Eliminate all unit productions)

- Eliminate all unit productions from the following grammar.

$$
\begin{aligned}
S &\rightarrow AB \mid CB \mid \mathbf{a}A \\
A &\rightarrow B \mid \mathbf{ab} \\
B &\rightarrow C \mid A\mathbf{b} \\
C &\rightarrow A \mid \mathbf{a}
\end{aligned}
$$

Outline

Eliminate all Useless Productions

Definition (Useful Variable)

A variable A is useful if there is at least one derivation

$$S \stackrel{*}{\Rightarrow} uAv \stackrel{*}{\Rightarrow} w$$

with $u, v \in (V \cup T)^*$ and $w \in T*$.

Definition (Useless Variable)

A variable is useless if it is not useful.

Definition (Useless Production)

A production $A \to u$ is useless if u contains a useless variable.

Eliminate all Useless Productions

- To determine which variables are useless,
  - Set $V' = \varnothing$.
  - Add to V' all variables A for which there is a production $A \to x_1 x_2 \cdots x_k$ with all $x_i \in V' \cup T$.
  - Repeat the previous step until no more symbols are added to V'.
  - For each $A \in V'$, use a dependency graph to determine whether there is a derivation $S \stackrel{*}{\Rightarrow} uAv$ for some $u, v \in (V \cup T)^*$.
  - If $A \notin V'$ or if $A \in V'$, but there is no derivation $S \stackrel{*}{\Rightarrow} uAv$, then A is useless.

Eliminate all Useless Productions

Example (Eliminate all Useless Productions)

- Eliminate all useless productions from the following grammar.

$$S \to AS \mid BB \mid SC \mid C \mid SD\mathbf{a}$$
$$A \to ABD \mid \mathbf{a}$$
$$B \to AD \mid \mathbf{b}D$$
$$C \to BD \mid C\mathbf{a} \mid S\mathbf{b} \mid \mathbf{b}$$
$$D \to \mathbf{a}B \mid D\mathbf{b} \mid AD$$

Eliminate all Useless Productions

Example (Eliminate all Useless Productions)

- Eliminate all λ-productions, unit productions, and useless productions from the following grammar.

$$S \to SAS \mid \mathbf{b}A\mathbf{a}$$
$$A \to \mathbf{a}S \mid S\mathbf{b} \mid \mathbf{ab} \mid S \mid \lambda$$

Outline

Assignment

Assignment

- Section 6.1 Exercises 6, 7, 8, 9, 10, 11, 17, 23, 24, 25, 26.