CS 5000: Lecture 11

Vladimir Kulyukin
Department of Computer Science
Utah State University

Outline

- Stack Machines and Context-Free Grammars
- Chomsky Normal Form

Stack Machine ➔ CFG

Theorem: if $M = (\Gamma, \Sigma, S, \delta)$ is a stack machine, then there is a CFG G such that $L(G) = L(M)$.

Proof Sketch: Assume that $\Gamma \cap \Sigma = \varnothing$. Construct $G = (\Gamma, \Sigma, S, P)$ so that whenever $t \in \delta(z, A)$, add the production $A \rightarrow zt$ to P.

Example: Stack Machine for $\{a^n b^n\}$

read	pop	push
a	S	S1
ε	S	ε
b	1	ε

Example: Stack Machine for $\{a^n b^n\}$

- Here is a CFG obtained from the stack machine:
  - $S \rightarrow aS1$
  - $S \rightarrow \varepsilon$
  - $1 \rightarrow b$
  - Note that 1 is a symbol of the stack alphabet in this grammar.

Example: Stack Machine for $\{a^n b^{2n}\}$

read	pop	push
a	S	Sbb
ε	S	ε
b	b	ε

First Attempt

- Here the first try:
  - S ➔ aSbb
  - S ➔ ε
  - b ➔ b
- Anything wrong with this grammar?

Second Attempt

- The previous grammar is not context-free, because it contains the production b ➔ b.
- In our construction we broke the assumption that the stack and input alphabets have nothing in common.
- Solution: rename the stack alphabet to separate it from the input alphabet.

Separating Γ and Σ

read	pop	push
a	S	SBB
ε	S	ε
b	B	ε

Second Attempt

- Here is our second attempt after the renaming:
  - S ➔ aSBB
  - S ➔ ε
  - B ➔ b
- This is a CFG that accepts the same language.

The Empty String Lemma

Let G be a CFG (V, Σ, S, P) where P contains rules of the form $A \to \varepsilon$ where $A \in V$. There is a CFG G' such that

1. $L(G) = L(G')$.
2. If $\varepsilon \notin L(G)$, then there are no productions of the form $A \to \varepsilon$ in G'.
3. If $\varepsilon \in L(G)$, then there is a single ε-production in G', $S' \to \varepsilon$, where S' is the start symbol of G' and S' does not occur on the right hand side of any production of G'.

Proof Sketch: Part 1

Suppose that $\varepsilon \notin L(G)$. For every $C \in V$,
consider all productions $C \rightarrow w, w \neq \varepsilon$.
Suppose consists of variables $A_1, \ldots, A_n, B_1, \ldots, B_m$,
where for $1 \leq i \leq k, A_i \Rightarrow^* \varepsilon$. Add to G the productions
$C \rightarrow w'$, where w' is obtained from w by deleting
0 or more occurences of A_i's. For example,
$C \rightarrow aA_1BA_2$ yields 4 productions:
$C \rightarrow aA_1BA_2; C \rightarrow aBA_2; C \rightarrow aA_1B; C \rightarrow aB$.
When this process is complete, remove all ε-productions
from G. Then remove all productions that have on the
right hand side variables that do not occur on the left
hand side of any production. The resulting grammar is G'.

Example

$$S \rightarrow ABDC$$

$$C \rightarrow RS$$

$$B \rightarrow \varepsilon$$

$$A \rightarrow a$$

$$D \rightarrow d$$

$$R \rightarrow \varepsilon$$

$$S \rightarrow d$$

Example: Step 1

$$S \to ABDC$$
$$S \to ADC$$
$$C \to RS$$
$$C \to S$$
$$B \to \varepsilon$$
$$A \to a$$
$$D \to d$$
$$R \to \varepsilon$$
$$S \to d$$

Example: Step 2

$$S \rightarrow ABDC$$

$$S \rightarrow ADC$$

$$C \rightarrow RS$$

$$C \rightarrow S$$

$$A \rightarrow a$$

$$D \rightarrow d$$

$$S \rightarrow d$$

Example: Step 3

$$S \rightarrow ADC$$
$$C \rightarrow S$$
$$A \rightarrow a$$
$$D \rightarrow d$$
$$S \rightarrow d$$

Proof Sketch: Part 2

If $\varepsilon \in L(G)$, then add a new start symbol to G', say S', and add two new productions: $S' \rightarrow S$ and $S' \rightarrow \varepsilon$. Then repeat the ε-rule elimination process of the previous step on all rules except these two ones.

A➔B Elimination Theorem

Theorem : Let G be a CFG, there is another CFG G' such that :

1. $L(G) = L(G')$;
2. If $\varepsilon \notin L(G)$, then there are no productions of the form $A \to \varepsilon$ in G'.
3. If $\varepsilon \in L(G)$, then there is a single ε - production in G' $S' \to \varepsilon$, where S' is the start symbol of G' and S' does not occur on the right hand side of any production in G'.
4. G' has no productions of the form $A \to B$, where $A \in V, B \in V$.

Proof Sketch

- By the previous theorem we may assume that conditions 2 and 3 are satisfied.
- To satisfy condition 4, construct a new production set *P'* by first
  - including into it all productions from *P* that are not of the form $A \rightarrow B$, where *A* and *B* are variables;
  - for each pair of variables *A* and *B*, if $A \rightarrow^* B$ according to *P*, then add in *P'* $A \rightarrow w$ if $B \rightarrow w$ is a production of *P* and *w* is not in *V*.

Normal Forms

- A normal form is a standard form of representing every object from a given set of objects.
- Example: S = { 10/20, 15/30, 14/28, 100/200, 45/90, 3/6 }.
- Every object in S can be represented as ½ by reducing it to lowest terms.

Chomsky Normal Form (CNF)

$G=\left(V,\Sigma,S,P\right)$ is in Chomsky Normal Form (CNF) iff each production in P has one of the following 3 forms:

1. $S\rightarrow\varepsilon.$
2. $A\rightarrow BC$, where $A,B,C\in V$.
3. $A\rightarrow a$, where $A\in V, a\in\Sigma$.
4. If $S\rightarrow\varepsilon\in P$, then $B\neq S, C\neq S$ in rule 2.

CNF Theorem

Theorem : Let G be a CFG. There is a CNF grammar G' such that $L(G) = L(G')$.

Proof Sketch : Assume that G satisfies conditions 2, 3, and 4 of the previous theorem. The set of productions P' is obtained from P as follows.

If $A \rightarrow xay$ is a production in P, where a is a terminal and x and y are strings of variables and terminals such that not both of them are empty, add the production $C_a \rightarrow a$ to P. Include all productions from P of the form $A \rightarrow a$ and $A \rightarrow BC$. For every production of the form $A \rightarrow B_1 ... B_m, m \geq 3$, add the following m-1 productions into P' :

$A \rightarrow B_1 C_1$

$C_1 \rightarrow B_2 C_2$

...

$C_{m-2} \rightarrow B_{m-1} B_m.$

Example: CNF Conversion

S → bA

S → aB

A → bAA

A → aS

A → a

B → aBB

B → bS

B → b

CNF Conversion: Stage 1

- $A \rightarrow a$
- $B \rightarrow b$
- Replace $S \rightarrow bA$ with $S \rightarrow C_bA$; $C_b \rightarrow b$
- Replace $S \rightarrow aB$ with $S \rightarrow C_aB$; $C_a \rightarrow a$
- Replace $A \rightarrow bAA$ with $A \rightarrow C_bAA$
- Replace $A \rightarrow aS$ with $A \rightarrow C_aS$; $C_a \rightarrow a$
- Replace $B \rightarrow aBB$ with $B \rightarrow C_aBB$
- Replace $B \rightarrow bS$ with $B \rightarrow C_bS$

CNF Conversion: Stage 2

- Replace $A \rightarrow C_bAA$ with
- $A \rightarrow C_bD_1$ and $D_1 \rightarrow AA$
- Replace $B \rightarrow C_aBB$
- $B \rightarrow C_a D_2$ and $D_2 \rightarrow BB$

CNF Conversion: Stage 3

- $S \rightarrow C_b A \mid C_a B$
- $A \rightarrow C_a S \mid C_b D_1 \mid a$
- $B \rightarrow C_b S \mid C_a D_2 \mid b$
- $D_1 \rightarrow AA$
- $D_2 \rightarrow BB$
- $C_a \rightarrow a$
- $C_b \rightarrow b$

Recommended Reading

- Section 10.3.

References

- Moll, R.N., Arbib, M.A., Kfoury, A.J. (1988). *An Introduction to Formal Language Theory*. Springer-Verlag.