Lecture 9: Properties of Context-Free Language

Instructor: Ketan Mulmuley
Scriber: Yuan Li

February 3, 2015

1 PDA and CFL continued

Let us continue of proof of equivalence between NPDAs and CFLs.

Theorem 1.1. *NPDAs accept precisely CFLs.*

Proof. In the last class, we have proved that for any CFL L, there exists an NPDA M such that $N(M) = L$. It remains to prove, for any given

$$M = (Q, \Sigma, \Gamma, \delta, q_0, z_0, \emptyset),$$

$N(M)$ is CFL. It suffices to construct a CFG G such that $L(G) = N(M)$. Let the grammar $G = (V, T, P, S)$, where $V = \{[q, A, p] : p, q \in Q, A \in \Gamma\} \cup \{S\}$, $T = \Sigma$. Grammar G will be such that $[q, A, p] \Rightarrow_G^* x$ if and only if $(q, x, A) \mapsto_M^* (p, \epsilon, \epsilon)$.

Let us add all production rules.

- For all $p \in Q$, add production $S \to [q_0, z_0, p]$.

- If $\delta(q, a, A)$, where $q \in Q$, $a \in \Sigma \cup \{\epsilon\}$, $A \in T$, contains

$$(q_1, B_1 B_2 \dots B_m),$$

  where $q_1 \in Q$ and $B_i \in \Gamma$, then for each $q_2, \dots, q_{m+1} \in Q$, add a production rule

$$[q, A, q_{m+1}] \to a[q_1, B_1, q_2][q_2, B_2, q_3] \dots [q_m, B_m, q_{m+1}].$$

  If $m = 0$, add production rule $[q, A, q_1] \to a$.

We leave it as an exercise to prove $L(G) = N(M)$. □

2 Pumping Lemma for CFL

Let $L = \{0^n 1^n : n \geq 1\}$, which is a CFL with the following grammar

$$S \to 01 \mid 0S1.$$

Let $L = \{a^i b^c c^i : i \geq 1\}$, and $\Sigma = \{a, b, c\}$. Is L a CFL? The answer is No. In order to prove it, we need a pumping lemma for CFL, similar with that for regular language.

Lemma 2.1 (Pumping Lemma for CFL)**.** *Let L be a CFL. There exists a constant n, depending only on L, such that if $z \in L$, $|z| \geq n$, then there exists $u, v, w, x, y \in \Sigma^*$ such that*

(1) $z = uvwxy$.

(2) $|vx| \geq 1$.

(3) $|vwx| \leq n$.

(4) $uv^i wx^i y \in L$ for all $i \geq 0$.

Proof. Fix G in Chomsky normal form such that $L(G) = L$. Let k be the number of variables in G, and let $n = 2^{k+1}$.

For any $z \in L$ with $|z| \geq n$, *any* derivation tree for z has depth at least $\log_2 |z| \geq k + 1$. (Recall that all production rules in Chomsky normal form are of the form $A \to a \mid BC$, where a is a terminal and B, C are variables. Thus the derivation tree of Chomsky normal form is a binary tree.) Then any parse tree for z mush have a path of length $\geq k + 1$.

Since the path has length at least $k + 1$, one variable must exist at least twice. Consider a node A on that path such that A appears again on the subpath, and the length of the subpath is at most $k + 1$. The desired strings u, v, w, x, y are illustrated in the Figure 1. Since the height of the subtree is at most $k + 1$, we have $|vwx| \leq 2^{k+1} = n$, which verifies (3). We claim v and x can not be empty simultaneously, because the production rule at root A is of the form $A \to BC$, and both B and C are not nullable. Thus $|vx| \geq 1$, which verifies (2). For (4), note that

$$A \to_G^* vAx \to_G^* v^2 Ax^2 \to_G^* v^3 Ax^3 \ldots$$

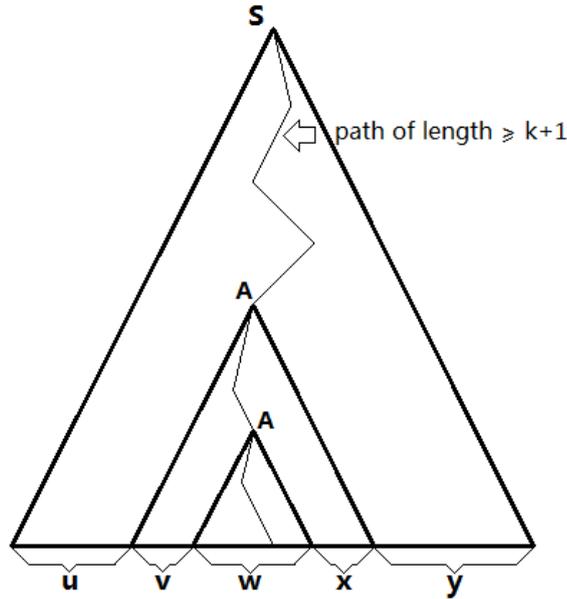and also $A \to w$. The proof is complete. $\square$

Figure 1: Parse tree for z

Now, let us apply pumping lemma to prove some languages are not CFL. By the way, we have seen pumping lemma for regular languages and CFL, which may look fancy at the first glance. However, when the story of real complexity theory begins, all these things fade.

Proposition 2.2. $L = \{a^i b^i c^i : i \geq 1\}$ *is not CFL.*

Proof. Suppose to the contrary L is CFL. Let n be as in the pumping lemma. Take $i = n$, and
$$z = a^i b^i c^i \in L.$$

By pumping lemma, there exists u, v, w, x, y such that $z = uvwxy$, $|vx| \geq 1$, $|vwx| \leq n$, and $uv^j wx^j y \in L$ for any $j \geq 0$. Let us do cases analysis. Since $|vwx| \leq n$, then vwx is either contained in one block (a-block, b-block or c-block), or two consecutive blocks (ab-block or bc-block). In any case, applying pumping lemma, $uv^2 wx^2 y \in L$ will have unequal number of a's, b's and c's, which is a contradiction. $\qquad\square$

Proposition 2.3. $L = \{a^i b^j c^i d^j : i, j \geq 1\}$ *is not CFL.*

Proof. Let n be as in the pumping lemma, and let

$$z = a^n b^n c^n d^n \in L.$$

By pumping lemma, there exists u, v, w, x, y such that $z = uvwxy$, $|vx| \geq 1$, $|vwx| \leq n$, and $uv^i wx^i y \in L$ for any $i \geq 0$. Since $|vwx| \leq n$, string vwx must be contained in one block, or two consecutive blocks. In either case, we will be able to pump z to get a contradiction, where the details are left to the reader. $\square$