Lec 12. Properties of PDA and Pumping Lemma for CFL

Eunjung Kim

Closure and non-closure properties of CFL

Context-free languages are closed under

- substitution
- union
- concatenation
- kleene star ($*$) and positive star ($+$)
- reversal
- intersection with a regular language

and not closed under

- intersection
- complementation
- $L_1 - L_2$

SUBSTITUTION

Given a CFL L over Σ and $a \in \Sigma$, we want to define a new language by substituting any occurrence of a by all strings of L_a. Here L_a is a CFL for each $a \in \Sigma$.

FORMAL DEFINITION OF SUBSTITUTION

For a finite alphabet Σ, let s be a mapping from Σ to the set of all languages, called a <u>subsitution</u> on Σ.

- For a string $w = a_1, \ldots, a_n \in \Sigma^*$, $s(w)$ is defined as

$$s(a_1) \cdot s(a_2) \cdot \cdots \cdot s(a_n).$$

- For a language L over Σ, $s(L)$ is defined as

$$\bigcup_{w \in L} s(w).$$

Substitution: example

- Let s be a substitution on $\Sigma = \{0, 1\}$ with $s(0) = \{a^n b^n : n \geq 1\}$ and $s(1) = \{aa, bb\}$.

- Let $L = \{0^i : i \geq 0\}$. Then $s(L)$ is the set of all strings of the form

  ?????????????????????????

- Let $L = \{10\}$. Then $s(L)$ is the set of all strings of the form

  ?????????????????????????

CFL's closed under substitution

Theorem

If L is a CFL and $s(a)$ is a CFL for each $a \in \Sigma$, then $s(L)$ is a CFL.

Union

Concatenation

Kleene and positive closure

Reversal

Intersection with a regular language

Theorem

If L is a CFL and R is a regular language, then $L \cap R$ is a CFL.

Intersection with a CFL

Not necessarily closed

The previous construction of PDA from a PDA and DFA extends to PDA with another PDA?

i.e. is $L_1 \cap L_2$ a CFL if both L_1 and L_2 are CFLs?

Intersection with a CFL

Not necessarily closed

The previous construction of PDA from a PDA and DFA extends to PDA with another PDA?

i.e. is $L_1 \cap L_2$ a CFL if both L_1 and L_2 are CFLs?

Not necessarily. Example.

- Consider $L_1 = \{a^n b^n c^i \mid n, i \geq 0\}$, $L_2 = \{a^i b^n c^n \mid n, i \geq 0\}$
- $L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 0\}$ is not CFL (via pumping lemma for CFL, next lecture).

Intersection with a CFL

Not necessarily closed

The previous construction of PDA from a PDA and DFA extends to PDA with another PDA?

i.e. is $L_1 \cap L_2$ a CFL if both L_1 and L_2 are CFLs?

Not necessarily. Example.

- Consider $L_1 = \{a^n b^n c^i \mid n, i \geq 0\}$, $L_2 = \{a^i b^n c^n \mid n, i \geq 0\}$
- $L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 0\}$ is not CFL (via pumping lemma for CFL, next lecture).

Why the product of two PDA is not PDA in general? You need two stacks, which makes it strictly more powerful.

Not necessarily closed with

Let L_1 and L_2 be context-free languages.

Not necessarily CFL

- $\bar{L_1} := \Sigma^* \setminus L_1$
- $L_1 - L_2$.

How about $L - R$, where L is CFL and R is regular?

Reminder: Pumping lemma for regular language

Pumping lemma: tool to prove nonregularity

Let A be a regular language. Then there exists a number p (called the pumping length) such that any string $w \in A$ of length at least p, w can be written as $w = xyz$ such that the following holds:

1. $|y| \geq 1$,
2. $|xy| \leq p$,
3. $xy^iz \in A$ for every $i \geq 0$.

Proof idea: DFA for A has a finite (constant) number of states.

Pumping lemma

Pumping lemma for CFL

Let A be a context-free language. Then there exists a number p (the pumping length) such that any string $w \in A$ of length at least p, w can be written as $w = uvxyz$ such that the following holds:

1. $|vy| \geq 1$,
2. $|vxy| \leq p$,
3. $uv^ixy^iz \in A$ for every $i \geq 0$.

Pumping lemma

Proof idea: For a sufficiently long string *w* and its parse tree, some variable is used at least twice.

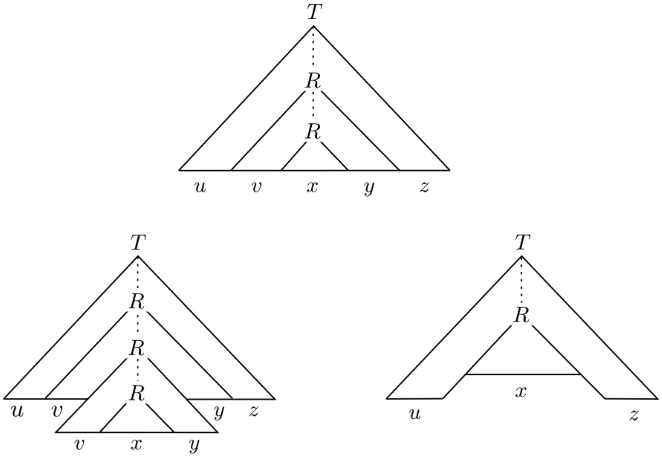

Figure 2.35 from Sipser 2012.

Pumping lemma for CFL, proof

There exists a context-free grammar $G = (V, \Sigma, S, R)$ with $L(M) = A$.

- Let b be the max number of symbols in the rhs of a rule.
- In any parse tree in this grammar, an internal node has $\leq b$ children.
- Any parse tree has $\leq b^h$ leaves, where h is the height of a parse tree in G.
- Let $p := b^{|V|+1}$.
- If $w \in A$ has length at least p, then its parse tree has $\geq p = b^{|V|+1}$ leaves, and height at least $|V| + 1$.
- Choose a parse tree τ yielding w with minimum number of nodes.
- Take a longest root-to-leaf path Q in τ; has length at least $|V| + 1$.
- Q has at least $|V| + 2$ nodes; only the last node is a terminal, the other $\geq |V| + 1$ nodes are variables.

Pumping lemma for CFL, proof

Let X be a variable which occurs twice in the <u>last</u> $|V| + 1$ variables on Q. Rewrite $w = uvxyz$, where vxy is the yield of the first X, and x is the yield of the second X.

1. $uv^ixy^iz \in A$: replacing the subtree rooted at the second X by the one rooted at the first X (or vice versa)
2. $|vy| \geq 1$: if $vy = \epsilon$, then replacing the subtree rooted at the first X by the subtree rooted at the second X leads to a parse tree with strictly smaller number of nodes. Contradicts the choice of τ.
3. $|vxy| \leq p$: the subtree rooted at the first X has height at most $|V| + 1$ by the choice of X. It has $\leq b^{|V|+1}$ leaves, thus its yield vxy has length $\leq b^{|V|+1} = p$.