Introduction to the Theory of Computation

XU Ming

School of Software Engineering, East China Normal University

June 4, 2023

OUTLINE

- Pumping Lemma for Context-Free Languages
- Applications of the Pumping Lemma for CFL's

Pumping Lemma for Context-Free Languages

Now, we'll develop tool for showing that certain language is not context-free.

The theorem, called the pumping lemma for context-free languages, says that in any sufficiently long string in a CFL, it is possible to find at most two short, nearby substrings, that we can "pump" in tandem. That is, we may repeat both of the string i times, for any integer i, and the resulting string will be still in the language.

This theorem is similar to the pumping lemma for regular languages, which says we can always find one small string to pump.

Parse tree of a CNF grammar is a binary tree, which has some convenient properties, one of which deals with the shape and size of the tree.

Theorem 7.9

Suppose that we have a parse tree according to a CNF grammar $G = (V, T, P, S)$, and suppose that the yield of the tree is a terminal string w. If the length of the longest path is n, then $|w| \leq 2^{n-1}$.

Proof The proof is based on a simple induction on n. □

Theorem 7.10

Let L be a CFL. Then there exists a constant n such that if z is any string in L such that $|z| > n$, then we can write $z = uvwxy$, subject to the following conditions:

1. $|vwx| \leq n$. *(That is, the middle portion is not too long.)*
2. $|vx| > 0$. *(That is, at least one of strings we pump must not be empty.)*
3. $\forall\, k \geq 0,\ uv^k wx^k y \in L$. *(That is, the two strings v and x may be "pumped" any number of times, including 0, the resulting string will still in L.)*

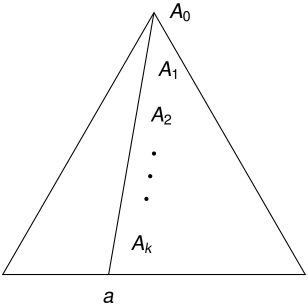
That is,

$$\exists\, n \in \mathbb{N} : \forall\, z \in L : \left[|z| > n \rightarrow \exists \text{ split } z = uvwxy : \left(\begin{array}{l} |vx| > 0 \ \wedge \\ |vwy| \leq n \ \wedge \\ \forall\, k \in \mathbb{Z}_{\geq 0} : uv^k wx^k y \in L \end{array} \right) \right.$$

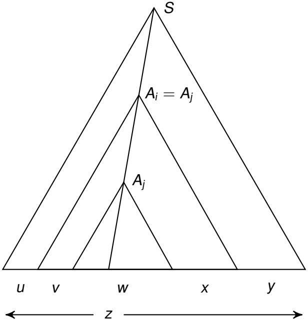
Proof Let $L \neq \emptyset$, $L \neq \{\epsilon\}$, we can find a CNF grammar G for L such that $L(G) = L \setminus \{\epsilon\}$. Let G have m variables and choose $n = 2^m - 1$. Suppose that z in L is of length greater than n.

By the Theorem 7.9, any parse tree whose longest path is of length m or less must have a yield of length n or less. Such a parse tree cannot have yield z, since it is too long. Thus, any parse tree with yield z has a path of length at least $m + 1$.

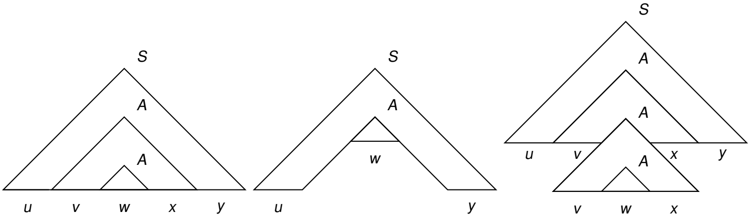
Now we have the longest path in the tree for z, there are at least $m+1$ occurrences of variables $S = A_0, A_1, \cdots, A_k$ ($k \geq m$) on the path.



As there are only m different variables in V, there are at least two variables $A_i = A_j$, where $k - m \leq i < j \leq k$. We divide the tree as



If $A_i = A_j = A$, then we can construct new parse trees from the original tree as follows. Thereby, we have proved the pumping lemma. □



Applications of the Pumping Lemma for CFL's

Applications of the Pumping Lemma for CFL's

Example

Consider language $L = \{0^m 1^m 2^m \mid m \geq 1\}$.

Suppose L is context-free. Then there is an integer n ensured by the pumping lemma. Pick $z = 0^n 1^n 2^n$ and break it as $z = uvwxy$, where $|vwx| \leq n$, and v and x are not both ϵ.

Then we know that vwx cannot involve both 0's and 2's, since the last 0 and the first 2 are separated by $n + 1$ positions. In following two cases, we'll find a contradiction, thus L is not a CFL.

1. vwx has no 2's. Then vx consists of only 0's and 1's, and has at least one of these symbols. By pumping lemma, we have $uwy \in L$, which however is impossible since it has n 2's, and fewer than n 0's or 1's, or both.
2. vwx has no 0's. The analysis is similar.

Example

Let $L = \{ww \mid w \in \{0,1\}^*\}$. Suppose the L is context-free. Then, let n be the pumping lemma constant. Consider the string $z = 0^n 1^n 0^n 1^n$.

Break $z = uvwxy$, with $|vwx| \leq n$ and $|vx| > 0$. We'll show that $uwy \notin L$, and thus show L not to be CFL, by contradiction.

Example

Prove language $L = \{0^i 1^{i^2} \mid i \in \mathbb{N}\}$ not to be context-free.

Let n be the pumping lemma constant and consider $z = 0^n 1^{n^2}$. We break $z = uvwxy$ according to the pumping lemma. If vwx consists only of 0's, the uwy has n^2 1's and fewer that n 0's; it is not in the language.

If vwx has only 1's, then we derive a contradiction similarly. If either v or x has both 0's and 1's, then uv^2wx^2y is not in 0^*1^*, and thus could not be in the language.

Finally, consider the case where v consists of 0's only, say k 0's, and x consists of m 1's only, where k and m are both positive.

Then for all i, $uv^{i+1}wx^{i+1}y$ consists of $n + ik$ 0's and $n^2 + im$ 1's. If the number of 1's is always to be the square of the number of 0's, we must have, for some positive k and m:

$$(n + ik)^2 = n^2 + im, \text{ or } 2ink + i^2k^2 = im.$$

But the left side grows quadratically in i, while the right side grows linearly, and so this equality for all i is impossible. We conclude that for at least some i, $uv^{i+1}wx^{i+1}y$ is not in the language and have thus derived a contradiction in all case.

Homework

Exercises 7.2.1(b)