Pumping Lemma for Context Free Language

Ayush Tandon (2K22/CO/133)

October 29, 2024

1 Introduction

In formal language theory, the Pumping Lemma is a fundamental in theoretical computer science for analyzing language properties. Pumping lemma for context free language (CFL) is used to prove that a language is not a context-free language.

2 Prerequisites

Context-Free Language:
A **CFL** is a language that can be generated by a **context-free grammar** (CFG), also known as a **Type 2 grammar** according to the Chomsky hierarchy. Context-Free Languages are characterized by productions in the form $A \to \alpha$, where A is a non-terminal symbol, and α is a string of terminals and/or non-terminals.

Pushdown Automata (PDA):
A **Pushdown Automata** is a theoretical machine that recognizes context-free languages. Unlike finite automata, a PDA has access to a stack, allowing it to process nested structures, which makes it suitable for parsing languages with recursive patterns. A language is context-free if and only if there exists a PDA that accepts it.

Chomsky Hierarchy:
The Chomsky hierarchy classifies formal languages into four types based on their generative power:

- **Type 0: Recursively Enumerable Languages** (accepted by a Turing machine)
- **Type 1: Context-Sensitive Languages** (accepted by a Linear-Bounded Automaton)
- **Type 2: Context-Free Languages** (accepted by a Pushdown Automaton)
- **Type 3: Regular Languages** (accepted by a Finite Automaton)

3 Definition

Let L be a context-free language. According to the Pumping Lemma for CFLs, there exists a constant p (known as the **pumping length**) such that any string $w \in L$ with $|w| \geq p$ can be decomposed into **five** substrings $w = uvxyz$ that satisfy the following conditions:

1. $|vxy| \leq p$
2. $|vy| \geq 1$ (i.e., either v or y is not empty)
3. For all $i \geq 0$, the string $uv^ixy^iz \in L$

These conditions imply that if L is context-free, then for any long enough string in L, it can be "pumped" by repeating portions of it, and the resulting strings will also belong to L.

4 Steps to Prove a Language is not Context-Free

To demonstrate that a language L is not context-free using the Pumping Lemma, follow these steps:

1. **Assume that L is context-free.**
2. **Identify the pumping length p.** According to the Pumping Lemma, there exists a pumping length p such that any string $w \in L$ with $|w| \geq p$ can be decomposed into five parts $w = uvxyz$ satisfying the lemma's conditions.
3. **Choose a string w in L with $|w| \geq n$.** Select a specific string w in the language L that is longer than or equal to the pumping length p.
4. **Divide w into substrings $uvxyz$.** Decompose w into five parts: $w = uvxyz$, where the substrings satisfy:
   - $|vxy| \leq p$
   - $|vy| \geq 1$
   - For any $k \geq 0$, $uv^kxy^kz \in L$
5. **Show that $uv^kxy^kz \notin L$ for some k.** Select a value of k, often $k = 2$, and demonstrate that the resulting string does not belong to L.
6. **Analyze all possible decompositions of w into $uvxyz$.** Consider all potential ways to divide w as $uvxyz$ and verify that none of these divisions satisfy all three pumping conditions simultaneously.
7. **Conclude that w cannot be pumped, leading to a contradiction.** Since w does not meet the conditions of the Pumping Lemma, this contradicts the assumption that L is context-free.
   Therefore, L is not a context-free language.

5 Illustration

5.1 Example 1

Consider the language $L = \{a^nb^nc^n \mid n \geq 1\}$. We want to determine whether L is context-free by using the Pumping Lemma.

Solution

Assume, for contradiction, that L is a context-free language. Then, according to the Pumping Lemma, there must exist a **pumping length** p such that any string $w \in L$ with $|w| \geq p$ can be split into five parts, $w = uvxyz$, such that the following conditions hold:

- $|vxy| \leq p$
- $|vy| \geq 1$ (at least one of v or y is non-empty)
- For all $k \geq 0$, $uv^kxy^kz \in L$.

Step 1: Choosing the string w Let $w = a^pb^pc^p$. Here, the string consists of p copies of a, followed by p copies of b, followed by p copies of c. Clearly, $|w| = 3p \geq p$, so this choice of w meets the requirement that $|w| \geq p$. We take p=4.

Step 2: Decomposing w as $uvxyz$ According to the Pumping Lemma, we can divide $w = a^4b^4c^4$ into five parts: $w = uvxyz$, where $|vxy| \leq p$ and $|vy| \geq 1$.

The condition $|vxy| \leq p$ implies that vxy must be within the first p characters of w, i.e., within the section of a's or at most into the section of b's. This restriction on vxy will be crucial in leading to a contradiction, as it limits v and y to contain only a's or b's, or at most a mix of both, without reaching the c's.

Step 3: Analyzing cases for v and y We now consider two possible cases for v and y based on where they fall within $w = a^4b^4c^4$:

Case 1: Both v and y contain only one type of symbol. Suppose that both v and y contain only a's or only b's. We choose in this manner:

$$u = a, \quad v = aa, \quad x = abbbbc, \quad y = c, \quad z = cc$$

Then, by applying the pumping condition with $i = 2$, we get:

$$\begin{aligned}
w' &= uv^2xy^2z \\
w' &= aaaaaabbbbccccc \\
w' &= a^6b^4c^5 \\
w' &\notin L
\end{aligned}$$

This results in a string with unequal number of a's , b's and c's. For instance, if v contains a's, then w' will have extra a's, breaking the balanced structure $a^nb^nc^n$ required by L. Similarly, if v contains b's, w' will have extra b's.

In both cases, w' does not follow the pattern $a^nb^nc^n$, so $w' \notin L$, which is a contradiction.

Case 2: Either v or y contains more than one kind of symbol. We choose in this manner:

$$u = aa, \quad v = aabb, \quad x = b, \quad y = b, \quad z = cccc$$

Then, by applying the pumping condition with $i = 2$, we get:

$$\begin{aligned}
w' &= uv^2xy^2z \\
w' &= aa \quad aabbaabb \quad b \quad bb \quad cccc \\
w' &= a^4b^2a^2b^2c^4 \\
w' &\notin L
\end{aligned}$$

In this case, pumping v and y will not be of form $a^nb^nc^n$, as w' will contain a mixture of a's and b's in the wrong order or an imbalance between the numbers of a's, b's, and c's.

As a result, this pumped string w' will not be in L, again leading to a contradiction.

Conclusion: Since in all possible cases, we arrive at a contradiction, we conclude that $L = \{a^nb^nc^n \mid n \geq 1\}$ cannot be context-free. Therefore, the language L is *not* **context-free.**

5.2 Example 2

Consider the language $L = \{ww \mid w \in \{0,1\}^*\}$. We want to determine whether L is context-free by using the Pumping Lemma.

Solution

Assume, for contradiction, that L is a context-free language. Then, according to the Pumping Lemma, there must exist a **pumping length** p such that any string $w \in L$ with $|w| \geq p$ can be split into five parts, $w = uvxyz$.

Now, let's choose a specific string $w \in L$ such that $|w| \geq p$:

$$w = 0^p 1^p 0^p 1^p$$

For simplicity, let $p = 5$, so our string becomes:

$$w = 0^5 1^5 0^5 1^5 = 00000\,11111\,00000\,11111$$

According to the Pumping Lemma, we should be able to split w into parts u, v, x, y, and z such that $w = uvxyz$ and $uv^k xy^k z \in L$ for all $k \geq 0$.

Since $|vxy| \leq p$, the substring vxy can only contain symbols from the first half, $0^5 1^5$ or the second half, $0^5 1^5$, but not both. This is due to the constraint on $|vxy| \leq p$. Consider the following cases:

Case 1: vxy lies entirely within the first half $0^5 1^5$. In this case, pumping v and y would affect only the number of 0's and 1's in the first half, but leave the second half $0^5 1^5$ unchanged. For instance, if we pump $k = 2$, we obtain:

$$uv^2 xy^2 z = 0^{5+|v|} 1^{5+|y|} 0^5 1^5$$

This string does not belong to L because it does not have the form ww — the first half no longer matches the second half.

Case 2: vxy lies entirely within the second half $0^5 1^5$. Similarly, if we pump v and y in the second half, we will alter the number of 0's and 1's in the second half only, which again breaks the symmetry required for $w \in L$. For example, if $k = 2$, we get:

$$uv^2 xy^2 z = 0^5 1^5 0^{5+|v|} 1^{5+|y|}$$

This string also does not belong to L, as the second half no longer matches the first half.

Conclusion: In both cases, the resulting string $uv^k xy^k z \notin L$ for $k \neq 1$, which contradicts the Pumping Lemma. Therefore, $L = \{ww \mid w \in \{0,1\}^*\}$ is **not** a context-free language.

6 Applications of the Pumping Lemma

The Pumping Lemma for CFLs is primarily used in theoretical computer science and formal language theory. Its main applications include:

- **Proving Non-Context-Freeness:** It helps in proving that a given language is not context-free by showing that any sufficiently long string cannot be decomposed in a way that satisfies the pumping conditions.
- **Language Classification:** By distinguishing context-free languages from other types, it aids in classifying languages and understanding their properties in relation to regular languages and context-sensitive languages.
- **Theoretical Analysis in Automata Theory:** The lemma is foundational in the study of pushdown automata and context-free grammars, providing insights into their limitations and behavior.