Cheatsheet - Context-Free Languages

Fabio Lama – fabio.lama@pm.me

1. Intro

Context-free grammar defines a set of rules for connecting strings together, which is another way of representing languages. It works recursively describing the structure of the strings.

2. Definition

A **context-free grammar** is a 4-tuple (V, Σ, R, S) where:

- **Variables**: a finite set of symbols, denoted V.
- **Terminals**: a finite set of letters, denoted by Σ, which is disjoint from V.
- **Rules**: a finite set of mappings, denoted by R, with each rule being a variable and a string of variables and terminals.
- **Start variable**: a member of V, denoted by S. It is usually the variable on the left-hand side of the top rule.

3. Generating Strings

1. Start from the **starting symbol**, read its rule.
2. Find a **variable** in the rule of the starting symbol and **replace it** with **a rule** of that variable.
3. **Repeat step 2** until there are no variables left.

A **derivation** is a sequence of substitutions in generating a string. There may be more than one rule for a variable. Then we can use the " $|$ " symbol to indicate "or".

For example:

$$S \rightarrow bSa \mid ba$$

Respectively:

$$S \rightarrow bSa$$

$$S \rightarrow ba$$

$$S \Rightarrow bSa \Rightarrow bbaa$$

$$S \Rightarrow bSa \Rightarrow bbSaa \Rightarrow bbbaaa$$

$$S \Rightarrow bSa \Rightarrow bbSaa \Rightarrow bbbSaaa \Rightarrow bbbbaaaa$$

We say u **derives** v, or $u \Rightarrow^* v$ if there is a derivation from u to v.

3.1. Example

$$S \rightarrow aS \mid T$$

$$T \rightarrow b \mid \varepsilon$$

4. Language of a Grammar

The language of grammar is all the strings that can be derived from the starting symbol using the rules of the grammar.

The formal definition is:

$$\text{If} \quad G = (V, \Sigma, R, S) \quad \text{then} \quad L(G) = \{w \in \Sigma^* \mid S \Rightarrow^* w\}$$

4.1. Example

We have the grammar G_2:

$$S \rightarrow aS \mid T$$

$$T \rightarrow b \mid \varepsilon$$

A few strings in $L(G_2)$: a, ab, b, ε, aa, aab, and **not** in $L(G_2)$: ba, abb, $aabb$.

We define it formally:

$$L(G_2) = a^* \cup a^*b = \{a^i b^j \mid 0 \leq i, 0 \leq j \leq 1\}$$

5. Converting from Regular Expressions

5.1. Example 1

Let's convert ab^* to a context-free language:

- b^* can be written as $U \to bU \mid \varepsilon$
- ab^* can be written as $S \to aU$

In other words:

$$S \to aU$$
$$U \to bU \mid \varepsilon$$

5.2. Example 2

Let's convert $ab^* \cup b^*$ to a context-free language:

- b^* can be written as $U \to bU \mid \varepsilon$
- ab^* can be written as $S \to aU$
- $\cup$ is just an "or", which can be written as $\mid$

In other words:

$$S \to aU \mid U$$
$$U \to bU \mid \varepsilon$$

5.3. Example 3

Let's convert $ab^+ \cup b^+b$ to a context-free language:

- b^+ can be written as $U \to bU \mid b$
- ab^+ can be written as $S \to aU$
- b^+b can be written as $S \to bU$
- $\cup$ is just an "or", which can be written as $\mid$

In other words:

$$S \to aU \mid bU$$
$$U \to bU \mid b$$

5.4. Example 4

Let's convert $\Sigma^* a \Sigma^*$, where $\Sigma = \{a, b\}$, to context-free language:

- Strings starting with a, $U \to aX$, $X \in \Sigma^*$
- Strings starting with b, $U \to bX$, $X \in \Sigma^*$
- Empty string, $U \to \varepsilon$

In other words:

$$S \to UaU$$
$$U \to aU|bU|\varepsilon$$

5.5. Example 5

Let's convert $\Sigma\Sigma\Sigma^+$, a binary string of at least three in length, to context-free language:

- $\Sigma^+ = (a \cup b)^+$ which can be written as $U \to aU|bU|a \mid b$
- $\Sigma\Sigma^+$ can be written as $V \to aU \mid bU$
- $\Sigma\Sigma\Sigma^+$ can be written as $S \to aV \mid bV$

In other words:

$$S \to aV \mid bV$$
$$V \to aU \mid bU$$
$$U \to aU|bU|a \mid b$$

Or, alternatively:

$$S \to aaU|abU|baU \mid bbU$$
$$U \to aU|bU|a \mid b$$

Last updated 2023-01-21 15:00:42 UTC