CS 208 : Automata Theory and Logic DATE

Lecture - 27

Topic: Intuition of Context Free Grammar from PDA

Scribed by: Deevyanshu Malu (22B1029)
Checked and compiled by:

1 More on PDA

Let us take an example of a PDA A=$(Q, \Sigma, \Gamma, q_0, Z_0, \delta, F)$ which accepts the language $\{0^n 1^n \mid n \geq 1\}$. Here,

- $Q = \{q_0, q_1, q_2\}$
- $\Sigma = \{0, 1\}$
- $\Gamma = \{Z_0, X\}$
- q_0 is the initial state
- Z_0 is the initial stack symbol
- $F = \phi$
- δ is the transition function

The PDA for the same is as follows:

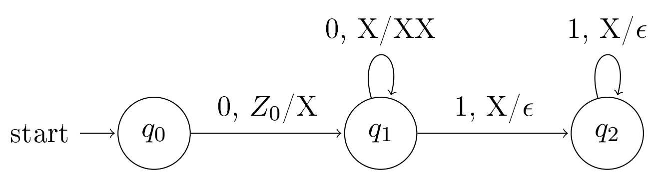


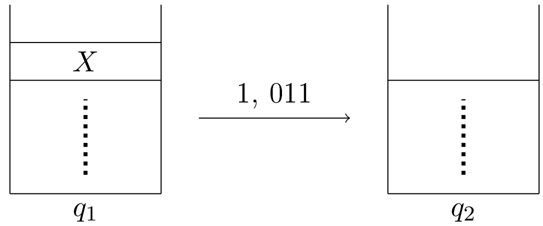
Figure 1: A PDA that accepts the language $\{0^n 1^n \mid n \geq 1\}$

1.1 Progress of stack

Suppose we take the word 0011 and run it on the PDA. The progress of the stack is as follows:

$$\begin{array}{ccccccccc} & & & & X & & & & \\ Z_0 & & X & & X & & X & & \\ q_0 & \xrightarrow{0} & q_1 & \xrightarrow{0} & q_1 & \xrightarrow{1} & q_2 & \xrightarrow{1} & q_2 \end{array}$$

We make the observation that to effectively pop one X from the stack and transition from q_1 to q_2, the word 1 or 011 must be used. The stack operation will be of the form:



Thus we see that both 1 and 011 serve the same function and thus we can replace one with the other and the resulting word will still be accepted.

$$w = 0011 \xrightarrow[\text{replaced by}]{\text{can be}} 000111 \text{ and } 000111 \xrightarrow[\text{replaced by}]{\text{can be}} 000011111$$

1.2 A new language

Let us define a language $L_{q_i Z_0 q_j}$ as the set of all strings that effectively pop Z_0 from the stack and transition from q_i to q_j. In the word $w = 0011$ we can say that:

- $0 \in L_{q_0 Z_0 q_1}$
- $0 \in L_{q_1 X q_1}$
- $1, 011 \in L_{q_1 X q_2}$
- $1 \in L_{q_2 X q_2}$
- $0011 \in L_{q_0 Z_0 q_2}$

The number of possible languages of the form $L_{q_i X q_j}$ is $n \times n \times k = n^2 k$, where n is the number of states and k is the number of stack symbols. This is because there are n states possible for q_i, n states possible for q_j and k stack symbols possible for X.

Now we can define the language accepted by empty stack as:

$$N(A) = L_{q_0 Z_0 q_0} \cup L_{q_0 Z_0 q_1} \cup L_{q_0 Z_0 q_2} \tag{1}$$

We will look at the language $L_{q_1 X q_2}$

- For the words with first letter 0, we can say that

  $$L_{q_1 X q_2} \supseteq 0 \cdot L_{q_1 X q_0} \cdot L_{q_0 X q_2} \cup 0 \cdot L_{q_1 X q_1} \cdot L_{q_1 X q_2} \cup 0 \cdot L_{q_1 X q_2} \cdot L_{q_2 X q_2} \tag{2}$$

  This is because the first 0 causes an X to be pushed onto the stack and state remains at q_1 and so now we have to pop X twice of the stack and transition to q_2 according to the definition of $L_{q_1 X q_2}$.

- For the words with first letter 1, we can say that

  $$L_{q_1 X q_2} \supseteq \{1\} \tag{3}$$

  This is because 1 causes the X to be popped from the stack and the transition to q_2, so nothing else needs to be done.

Thus in general, we can say that

$$L_{q_1 X q_2} \supseteq 0 \cdot L_{q_1 X q_0} \cdot L_{q_0 X q_2} \cup 0 \cdot L_{q_1 X q_1} \cdot L_{q_1 X q_2} \cup 0 \cdot L_{q_1 X q_2} \cdot L_{q_2 X q_2} \cup \{1\} \tag{4}$$

1.3 Context Free Grammar

Equation 4 implies that wherever we see a word of the language $L_{q_1Xq_2}$, we can replace it with any word from the languages on the RHS. Thus we can write:

$$L_1 \rightarrow 0L_2L_3 \mid 0L_4L_1 \mid 0L_1L_5 \mid 1$$

where $L_1 = L_{q_1Xq_2}$, $L_2 = L_{q_1Xq_0}$, $L_3 = L_{q_0Xq_2}$, $L_4 = L_{q_1Xq_1}$ and $L_5 = L_{q_2Xq_2}$.
Similarly, we can do the same for the languages $L_i \forall i \in \{1, 2, \ldots, n^2k\}$. This will form a **context free grammar**.

WIth proper reductions, we can say that the context free grammar generated for the language $L = \{0^n1^n \mid n \geq 1\}$ is:

$$S \rightarrow 0S1 \mid 01$$

We observe that the universal language $\Sigma^* = L((0+1)^*)$ also satisfies the context free grammar, but the language $L = \{0^n1^n \mid n \geq 1\}$ is the minimal language that satisfies the context free grammar.