4CS4-6:Theory of Computation(Pumping Lemma)

Lecture 09: Feb. 11, 2019

Prof. K.R. Chowdhary *: Professor of CS*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

9.1 Kleene Star properties of Regular Languages

The regular languages have infinitely many strings, which in turn consists of large number of substrings resulting from the Kleene Star operations. The name regular indicates the regularity in length. This fact can be used for reasoning about the languages. The approach taken is based on the fact that long strings going through the finite-state machines cannot avoid revisiting the states. Hence, if such strings go from the start state to a final state, one can traverse a loop described by the re-visitation, arbitrary number of times. And the corresponding automaton recognizing these substrings have cycles of states visited, representing Kleene Star operations in the regular expressions.

9.1.1 Pumping Lemma

The following theorem is based on the Kleene Star operations in regular expressions. When expressed rigorously, this idea forms the basis to show that certain languages are not regular, and takes the curious name *Pumping Lemma*.

Theorem 9.1 *Pumping Lemma: Let Q be the set of states in a finite automata M with $|Q| = n$, L be a regular set such that $w \in L$ and $|w| \geq n$. Having these conditions, w can be considered as made of substrings xyz, where $|xy| \leq n$ (because in x and y there are unique states)*[1]*, $|y| \geq 1$, then for all $k \geq 0$, $xy^kz \in L$.*

Pumping lemma is a tool to say about non-regularity in certain languages. Algorithms can be designed based on this concept to know whether a language accepted by finite automata is finite or infinite.

Let there be a finite automaton $M = (Q, \Sigma, \delta, q_0, F)$, and $|Q| = n$. Assume that a string $a_1a_2...a_m$, where $m \geq n$ is input to this FA. Due to this, the FA goes through the sequence of states $q_0, q_1, q_2, ..., q_m$. Since the number of states in this sequence, i.e., $m + 1$, is more than n (the total number of states in the FA), according to the *pigeonhole principle* at least one state has been visited twice. Let this state be $q_i = q_j$ in the sequence of states traversed, where $0 \leq i < j \leq m$, shown in Fig. 9.1. In the sequence of integers 0 to m, the string

[1]Since in x and y path there are unique states, and $|Q| = n$, therefore the unique states in xy are not more than Q. Hence, $|xy| \leq |Q|$.

$a_{i+1}a_{i+2}...a_j$ has length of at least one. Since $0 \leq i$ and $j \leq m$, the length of this string can be maximum m. Let $x = a_1a_2...a_i$, $y = a_{i+1}a_{i+2}...a_j$, and $z = a_{j+1}a_{j+2}...a_m$.

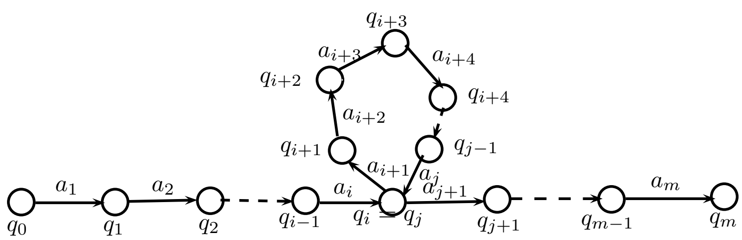

Figure 9.1: Principle of Pumping Lemma

If $q_m \in F$, then the string xyz is accepted by the FA. At the same time, another string xz is also accepted by the FA as follows:

$$\begin{aligned} \delta^*(q_0, xz) &= \delta^*(\delta^*(q_0, x), z) \\ &= \delta^*(q_i, z) \\ &= \delta^*(q_j, z), \text{ since } q_i = q_j \\ &= q_m \end{aligned}$$

It can be easily concluded that the finite automaton M (Fig. **??**) recognizes following strings:

1. $xyz \in L(M)$
2. $xz \in L(M)$

The first of above expressed as $xyz \in L(M)$ and can be generalized as

$$xy^kz \in L(M), \text{ for } k = 1 \tag{9.1}$$

For $k = 0$, equation 9.1 results to string xz. This means that a string $w = xyz$ having a substring y, if y is repeated any number of times, the total string is still be accepted by the FA. In other words, the substring enclosed between parenthesis (i.e. $y = a_{i+1}...a_j$) can be "pumped" any number of times, such that the states corresponding to the path of the substring forms a loop in the transition graph, and the resulting string $w = xy^kz$, with $k \geq 0$, is still accepted by the FA. Hence, the name "pumping theorem" or "pumping lemma" is given for this conclusion.

9.1.2 Examples on Pumping lemma

Example 9.2 *The language $L = \{a^nb^n \mid n \geq 0\}$ is not regular. Prove this statement using pumping lemma.*

Assume that L is regular and pumping lemma hold true for this. Let us consider some integers n, and let $w = a^nb^n \in L$. Using pumping lemma decompose w into xyz, such that $|xy| \leq n$, and $|y| \geq 1$ and for some $k > 0$, $y = a^k$. We note that, there is $xz = a^{n-k}b^n \notin L$. Since it contradicts the assumption, L is not regular.

Example 9.3 *Prove that language $L = \{a^n \mid n \text{ is prime}\}$, is not regular.*

Let the finite automaton corresponding to language L be M, and a string w is input to M, where $w = xy^n z$. Let $x = a^p$, $y = a^q$ and $z = a^r$, where $q > 0$ and $p \geq 0$, $r \geq 0$. Now, let us apply the pumping lemma for $w = xy^k z$, for $k \geq 0$.

$$\begin{aligned} w &= xy^k z \\ &= a^p (a^q)^k a^r \\ &= a^p\ a^{kq}\ a^r \\ &= a^{p+kq+r}. \end{aligned}$$

Hence, the value $n = p + kq + r$ should be a prime number with $k \geq 0$.

We prepare a strategy to counter it and try to show that $p + nq + r$ is not a prime. Let $k = p + 2q + r + 2$. Then,

$$\begin{aligned} p + kq + r &= p + (p + 2q + r + 2)q + r \\ &= p + pq + 2q^2 + qr + 2q + r \\ &= (p + 2q + r) + q(p + 2q + r) \end{aligned}$$
$$= (1 + q)(p + 2q + r)$$

Since it is factorable, $p + kq + r$ cannot be a prime. Since, it contradicts the assumption, L is not regular.

Example 9.4 *Show that $L = \{a^{n^2} \mid n \text{ is integer, and } n \geq 1\}$, is not regular.*

For n to be integer and $n \geq 1$, L consists all the strings whose length is perfect square. That is, $L = \{a^1, a^2, a^9, ...\}$. To begin with, we assume that L is regular. Let there is a string $w \in L$, where $w = xyz = a^{n^2}$, where n is integer in pumping lemma, indicating the number of states. Using the pumping lemma a^{n^2} can be written as xyz, where $1 \leq |y| \leq n$, and $xy^k z \in L$ for $k \geq 0$. Now, we try to prepare as a strategy that $|xy^k z|$ is not a perfect square.

For this we want to represent a^{n^2} in the form of $xy^k z$, and try to show that when original expression is represented as $xy^k z$ it fails to satisfy the criteria of a^{n^2}.

Since $|y|$ must be greater or equal to 1, let $|y| = m \geq 1$. Since we have assumed that w is perfect square, let there is some integer n, for which $|w| = n^2 = |xz| + |y| = (n^2 - m) + m$.

Now, let $k = 2$. Hence, $xy^2 z$ should be in L. Therefore,

$$|xy^2 z| = |xz| + |y^2| = (n^2 - m) + 2m = n^2 + m.$$

This m can be at the most n, which happens when all the states are in y. Hence, $n^2 + m \leq n^2 + n$. Therefore,

$$n^2 < |xy^2z| \leq n^2 + n < n^2 + 2n + 1 = (n+1)^2.$$

This indicates that the length of xy^2z lies between n^2 and $(n+1)^2$, hence the length is not a perfect square. This shows that $xy^2z \notin L$. Since this is in contradiction to our assumption, L is not regular.