Regular and Nonregular Languages: The Pumping Lemma for Regular Languages (Part 2)

COMS3003A: Lecture Note 8

April 22, 2021

Recap: The pumping lemma for regular languages

A weaker form of the above theorem, commonly known as *the* pumping lemma:

Theorem:
Suppose L is a regular language. Then there is an integer n so that for any $x \in L$ with $|x| \geq n$, there are strings u, v, and w so that

$$x = uvw \ ,$$

$$|uv| \leq n \ ,$$

$$|v| > 0 \ ,$$

$$\text{for any } m \geq 0, uv^m w \in L \ .$$

The correct argument can be visualized as a game we play against an opponent.

Our goal is to win the game by establishing a contradiction of the pumping lemma, while the opponent tries to foil us.

There are four moves in the game.

1. The opponent picks n.
2. Given n, we pick a string $x \in L$ of length equal or greater than n. We are free to choose any x, subject to $x \in L$ and $|x| \geq n$.
3. The opponent chooses the decomposition uvw, subject to $|uv| \leq n$ and $|v| > 0$. We have to assume that the opponent makes the choice that will make it hardest for us to win the game.
4. We try to pick m in such a way that the pumped string $uv^m w$ is not in L. If we can do so, we win the game.

A strategy that allows us to win whatever the opponent chooses is a proof that the language is not regular.

In this respect, Step 2 is crucial.
While we cannot force the opponent to pick a particular decomposition of x, we may be able to choose x so that the opponent is very restricted in Step 3, forcing a choice of u, v, and w that allows us to produce a violation of the pumping lemma on our next move.

Let $\Sigma = \{a, b\}$. Show that

$$L = \left\{ ww^R \mid w \in \Sigma^* \right\} \quad ,$$

where w^R denotes the reverse of w, is not regular.

Whatever n the opponent picks in Step 1, we can always choose a

$$x = a^n b^n b^n a^n \quad .$$

Because of this choice, and the requirement that $|uv| \leq n$, the opponent is restricted in Step 3 to choosing a v that consists entirely of a's.

In Step 4, we use $m = 0$.

The string obtained in this way has fewer a's on the left than on the right and so cannot be of the form ww^R.

Therefore L is not regular.

If we had chosen x too short, then the opponent could have chosen a v with an even number of b's.
In that case, we couldn't have reached a violation of the pumping lemma in Step 4.

Suppose we chose a string consisting of all a's, say

$$x = a^{2n} \ .$$

The opponent need only pick

$$v = aa \ .$$

Then $uv^m w \in L$ for every $m \geq 0$, and we lose.
Note that, if the opponent picked

$$v = a \ ,$$

then $uv^0 w \notin L$, but we *cannot* claim that we have reached a contradiction just because the pumping lemma is violated for some specific values of n or uvw.

Weak form of the pumping lemma

Sometimes we can get by without using n:
Suppose L is an infinite regular language. Then there are strings u, v, and w so that

$$|v| > 0 \text{ and}$$

$$uv^m w \in L \text{ for every } m \geq 0$$

Sketch of proof:
L has infinitely many elements.
Therefore, no matter how big n is, L contains an x with $|x| \geq n$.

This is sufficient to prove that

$$L = \left\{0^i 1^i \mid i \geq 0\right\}$$

is not regular.

- Suppose $v = 0^j$, for some $j \geq 1$: If $uv^{m'}w \in L$ for some $m' \geq 0$, then $uv^{m'+1}w \notin L$.
- Suppose $v = 1^j$, for some $j \geq 1$: Analogous to previous case.
- Suppose v contains 01: Then $uv^2w \notin L$.

The weak form is not sufficient in the case of the following two languages.

- $L = \left\{0^i y \mid i \geq 0, y \in \{0,1\}^* \text{ and } |y| \leq i\right\}$
  Choose $v = 0$. Then $0v^m 1 \in L$ for all $m \geq 0$.
- L_{pal}, the language of all palindromes over $\{0,1\}$.
  Choose $v = 0$. Then $1v^m 1 \in L_{\text{pal}}$ for all $m \geq 0$.

Even weaker form of the pumping lemma:

Suppose L is an infinite regular language.
Then there are integers p and q, with $q > 0$, so that for every $m \geq 0$, L contains a string of length $p + mq$.

In other words, the set of integers

$$\text{lengths}(L) = \{|x| \mid x \in L\}$$

contains the 'arithmetic progression' of all integers $p + mq$, where $m \geq 0$.

Proof
In the previous theorem, let $p = |u| + |w|$ and $q = |v|$.

This form is not sufficient to prove that

$$L = \{0^i 1^i \mid i \geq 0\}$$

is not regular, since

$$\text{lengths}(L) = \{0, 2, 4, \ldots\}$$

and thus contains all integers $0 + m2$, $m \geq 0$.
It is sufficient in the case of

- $\{a^{2^n} \mid \ n \geq 1\}$,
- $\left\{a^{n^2} \mid \ n \geq 1\right\}$ and
- $\{0^n \mid \ n \text{ prime}\}$

Consider

$$L = \{0^n \mid \ n \text{ prime}\} = \{0^2, 0^3, 0^5, 0^7, \ldots\}$$

We need to show that

- L is not regular, therefore that
- the set of primes cannot contain a set $\{p + mq \mid m \geq 0\}$, therefore that
- for any $p \geq 0$ and $q > 0$, there is an integer m so that $p + mq$ is not prime.

Let $m = p + 2q + 2$. Then

$$\begin{aligned} p + mq &= p + (p + 2q + 2)\,q \\ &= (p + 2q) + (p + 2q)\,q \\ &= (p + 2q)\,(1 + q) \end{aligned}$$

This is not prime.
This example shows that an FA is not powerful enough to determine for an arbitrary integer whether it is prime.

The pumping lemma cannot show a language is regular
Consider

$$L = \left\{a^i b^j c^j \mid i \geq 1 \text{ and } j \geq 0\right\} \cup \left\{b^j c^k \mid j, k \geq 0\right\}$$

L is not regular, but the conclusions of the pumping lemma hold.

Using the Myhill-Nerode theorem, we show that L is not regular.
Consider ab^j and ab^k, where $0 < j < k$. Then c^j distinguishes the two strings.
Then all equivalence classes $\left[ab^j\right]$, $j \geq 0$, are distinct. Therefore L is not regular.

The conclusions of the pumping lemma hold.
Take $n = 1$. Suppose $x \in L$ and $|x| \geq n$.
x can have two forms.

- $x = a^i b^j c^j$, where $i > 0$.
  Let $u = \lambda$, $v = a$, $w = a^{i-1} b^j c^j$.
  Then for any $m \geq 0$, $uv^m w = a^m a^{i-1} b^j c^j = a^{m+i-1} b^j c^j \in L$.
- $x = b^i c^j$.
  Let $u = \lambda$, and v equal the first symbol in x.
  Then for any $m \geq 0$, $uv^m w$ has m copies of the first symbol in x, and is in L.