CSE2315 Slides week 3

Matthijs Spaan Stefan Hugtenburg

Algorithmics group

Delft University of Technology

26 February 2020

This lecture

- Pigeonhole principle
- Nonregular languages
- Pumping lemma for regular languages

Riddle

(see also video Pumping Lemma (for Regular Languages))

Given that a human being can have at most 200,000 hairs on its head and that the population of New York is 18,976,457. What is the chance that two New Yorkers have exactly the same number of hairs on their head?

(a) 0
(b) small but nonzero
(c) about a half
(d) great but not 1
(e) 1

Pigeonhole Principle

Informally: If you put n objects into m holes and $n > m$, then there is at least one hole with 2 or more objects.

Formally: Let A and B be *finite* sets with n and m elements, with $n > m$, and let $f : A \to B$ be a function from A to B.
Then there are elements $x, y \in A$ with $x \neq y$ such that $f(x) = f(y)$.

Well-known nonregular language

A well-known nonregular language is:

$$L = \{a^n b^n \mid n \geq 0\}.$$

Intuition: A finite automaton cannot remember that it has read n as.

Pumping lemma: informally

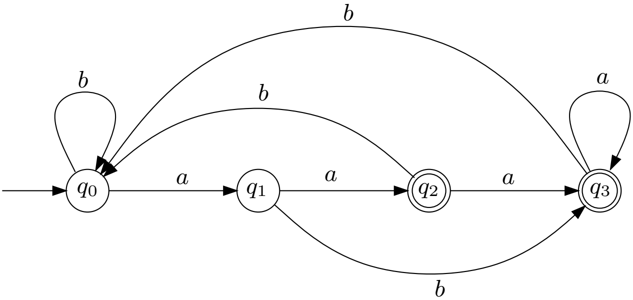
Let M be a given finite automaton with p states.
Let an input word w for M be given as well, with $|w| \geq p$.
What happens while M processes w?

M starts in q_0, the start state, and w must be read.
There are now $p - 1$ states "left" (differing from each other and from q_0) to read at least p letters of w.

Conclusion: There must be a state that is visited at least twice.

Exercise

Suppose we have the following DFA.



Which states are visited (at least) twice if the following words are read:

(a) aaa (b) $abaa$ (c) $abbaaa$

Pumping lemma for regular languages

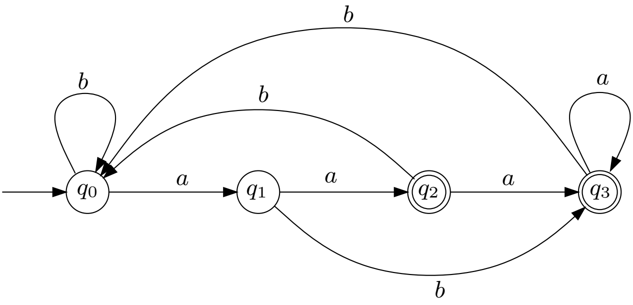
Theorem: Let L be a regular language.

- Then there exists some $p > 0$, the *pumping length of* L, and
- for every word $w \in L$ with $|w| \geq p$
- there exists a division of w into pieces xyz with $|xy| \leq p$ and $|y| > 0$, such that
- for all $i \geq 0$: $xy^iz \in L$.

Structure of quantifiers in the theorem: $\exists\forall\exists\forall$.

Exercise

For the regular language L recognized by the following DFA, we can take a pumping length $p = 4$:



For example, the word $w = abbaaa$ can be divided up as follows:

$x = a$, $y = bba$, and $z = aa$.

Verify the requirements of the theorem in this case.

Solution

Indeed, we have:

- $|w| = |abbaaa| = 6 \geq 4 = p$
- $|xy| = |abba| = 4 \leq 4 = p$
- $|y| = |bba| = 3 > 0$
- For all $i \geq 0$: $xy^iz = a(bba)^iaa \in L$. When reading this word, the machine 'loops' i through the states q_1, q_3, q_0 and back to q_1 when accepting the ith bba.

Use of the pumping lemma

The pumping lemma can be used to show that a language is not regular.

Remember: The pumping lemma CANNOT be used to show that a language *is* regular!

A game for two players

Setting: Player A claims L is not regular, player B contests this. How might this game proceed?

A: L is not regular!
B: L is regular! I say p is the pumping length of L.
A: Here is a word $w \in L$ with $|w| \geq p$. Now you need to give me a suitable division of this w.
B: Ok, here it is: $w = xyz$. (*)
A: Then indeed $|xy| \leq p$ and $|y| > 0$. But now I'm winning, because I found an i for which $xy^iz \notin L$. Therefore, you've lost! L is not regular!

(*) Since B wants to win, it has undoubtedly thought up the hardest division.

The division of w

In practice the game is obviously not played. It does help give a better understanding of the theorem, however.

If A claims, based on the pumping lemma, that L is not regular, then it must have a winning strategy:

1. If B claims to have found a pumping length ($\exists$), A can pick a word ($\forall$).
2. If B claims to have found a division ($\exists$), A can choose an i ($\forall$) such that $xy^iz \notin L$.

Apparently A can find an i for every division of $w = xyz$ by B such that $xy^iz \notin L$.

Scheme of usage of pumping lemma

Prove: Language L is not regular.
Proof method: By contradiction.

- Suppose L is regular.
  - The pumping lemma holds for L, so there exists a pumping length p.
  - Form a word w such that $|w| \geq p$ and $w \in L$.
  - Take into account every possibility of x, y and z with $w = xyz$ such that $|xy| \leq p$ and $|y| > 0$.
  - Show for each possibility that there exists an $i \geq 0$ such that $xy^iz \notin L$.
- Contradiction. Therefore, L is not regular.

Example

Suppose we have a language

$$L = \{a^n b^n \mid n \geq 0\}.$$

Use the pumping lemma to show that L is not regular.

(Solution: see Sipser Example 1.73)

Exercise

Let $\Sigma = \{a, b\}$ be an alphabet. Suppose we have a language

$$L = \{w \in \Sigma^* \mid n_a(w) = n_b(w)\}.$$

Show that L is not regular.

Solution

1. Suppose $L = \{w \in \{a, b\}^* \mid n_a(w) = n_b(w)\}$ is regular.
2. Let p be the pumping length of L according to the pumping lemma.
3. Take the word $w = a^p b^p$. Note that $|w| \geq p$ and $w \in L$.
4. Since $|w| \geq p$ and $w \in L$, we have that $y = a^n$ for some $n \geq 1$ and $n \leq p$.
5. According to the pumping lemma, we should have $xy^2z \in L$.
6. However, $n_a(xy^2z) = n_a(xyz) + n$, while $n_b(xy^2z) = n_b(xyz)$.
7. Therefore, $n_a(xy^2z) \neq n_b(xy^2z)$ and $xy^2z \notin L$.
8. Contradiction! So L is not regular.

Exercise

Let $\Sigma = \{a, b\}$ be an alphabet. Is the language

$$L' = \{w \in \Sigma^* \mid n_a(w) \neq n_b(w)\}$$

regular?

Hint: Remember the closure properties!

Answer

No, L' is not regular, since the complement of a nonregular language is nonregular.

If we suppose L' were regular, then $\overline{L'}$, the complement of L', must be regular too. Since $\overline{L'} = L$ with L the nonregular language from the previous exercise, $\overline{L'}$ is not regular either. Contradiction, so L' is not regular.

Remarks when using the pumping lemma

1. Often a division by cases is necessary, so consider all possible divisions of the chosen word.

   For example: $\{(ab)^n a^m \mid n < m\}$.
2. "Pumping down" (taking $i = 0$)

   For example: $\{a^n b^m \mid n > m\}$.
3. Use the closure properties of regular languages

   For example: $\{w \in \{a,b\}^* \mid n_a(w) \neq n_b(w)\}$.