CIS 770: Formal Language Theory

Pavithra Prabhakar

Kansas State University

Spring 2015

Finite Languages

Definition

A language is finite if it has finitely many strings.

Example

$\{0, 1, 00, 10\}$ is a finite language, however, $(00 \cup 11)^*$ is not.

Finiteness and Regularity

Proposition

If L is finite then L is regular.

Proof.

Let $L = \{w_1, w_2, \ldots w_n\}$. Then $R = w_1 \cup w_2 \cup \cdots \cup w_n$ is a regular expression defining L. $\square$

Are all languages regular?

Proposition

The language $L_{\mathrm{eq}} = \{w \in \{0,1\}^* \mid w \textit{ has an equal number of 0s and 1s}\}$ *is not regular.*

Proof?

No DFA has enough states to keep track of the number of 0s and 1s it might see. □

Above is a weak argument because $E = \{w \in \{0,1\}^* \mid w \text{ has an equal number of 01 and 10 substrings}\}$ is regular!

Proving Non-Regularity

Proposition

The language $L_{\mathrm{eq}} = \{w \in \{0,1\}^* \mid w \text{ has an equal number of 0s and 1s}\}$ *is not regular.*

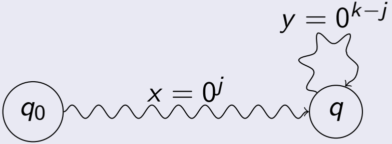
Proof.

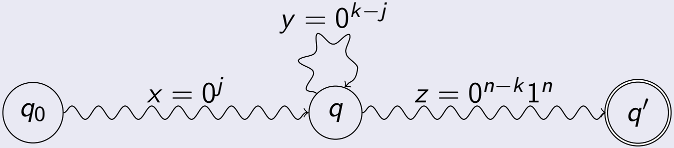
Suppose (for contradiction) L_{eq} is recognized by DFA $M = (Q, \{0,1\}, \delta, q_0, F)$, where $|Q| = n$.

- There must be $j < k \leq n$ such that $\hat{\delta}(q_0, 0^j) = \hat{\delta}(q_0, 0^k)$ ($= q$ say).
- Let $x = 0^j$, $y = 0^{k-j}$, and $z = 0^{n-k}1^n$; so $xyz = 0^n1^n$. $\cdot\cdot\!\longrightarrow$

Proving Non-Regularity

Proof (contd).





- We have $\hat{\delta}(q_0, 0^j) = \hat{\delta}(q_0, 0^k) = q$
- Since $0^n 1^n \in L_{\mathrm{eq}}$, $\hat{\delta}(q_0, 0^n 1^n) \in F$.

Pumping Lemma: Overview

Pumping Lemma

The lemma generalizes this argument. Gives the template of an argument that can be used to easily prove that many languages are non-regular.

Lemma

If L is regular then there is a number p (the pumping length) such that $\forall w \in L$ with $|w| \geq p$, $\exists x, y, z \in \Sigma^*$ such that $w = xyz$ and

1. $|y| > 0$
2. $|xy| \leq p$
3. $\forall i \geq 0.\ xy^i z \in L$

Proving the Pumping Lemma

Proof.

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA such that $L(M) = L$ and let $p = |Q|$. Let $w = w_1 w_2 \cdots w_n \in L$ be such that $n \geq p$. For $1 \leq i \leq n$, let $s_i = \hat{\delta}(q_0, w_1 \cdots w_i)$; define $s_0 = q_0$.

- Since $s_0, s_1, \ldots, s_i, \ldots s_p$ are $p + 1$ states, there must be j, k, $0 \leq j < k \leq p$ such that $s_j = s_k$ ($= q$ say).
- Take $x = w_1 \cdots w_j$, $y = w_{j+1} \cdots w_k$, and $z = w_{k+1} \cdots w_n$
- Observe that since $j < k \leq p$, we have $|xy| \leq p$ and $|y| > 0$.

$\cdot\cdot\!\longrightarrow$

Proof . . .

Technical Claim

Claim

For all $i \geq 1$, $\hat{\delta}(q_0, xy^i) = \hat{\delta}(q_0, x)$.

Proof.

We will prove it by induction on i.

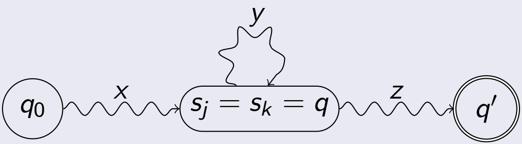
- Base Case: By our assumption that $s_j = s_k$ and the definition of x and y, we have $\hat{\delta}(q_0, xy) = s_k = s_j = \hat{\delta}(q_0, x)$.
- Induction Step: We have

$$\begin{aligned}\hat{\delta}(q_0, xy^{\ell+1}) &= \hat{\delta}(\hat{\delta}(q_0, xy^{\ell}), y)\\ &= \hat{\delta}(\hat{\delta}(q_0, x), y)\\ &= \hat{\delta}(q_0, xy) = \hat{\delta}(q_0, x)\end{aligned}$$

$\square$

Completing the Proof

Proof (contd).



- We have $\hat{\delta}(q_0, xy^i) = \hat{\delta}(q_0, x)$ for all $i \geq 1$
- Since $w \in L$, we have $\hat{\delta}(q_0, w) = \hat{\delta}(q_0, xyz) \in F$
- Observe, $\hat{\delta}(q_0, xz) = \hat{\delta}(\hat{\delta}(q_0, x), z) = \hat{\delta}(\hat{\delta}(q_0, xy), z) = \hat{\delta}(q_0, w)$. So $xz \in L$
- Similarly, $\hat{\delta}(q_0, xy^iz) = \hat{\delta}(\hat{\delta}(q_0, xy^i), z) = \hat{\delta}(\hat{\delta}(q_0, x), z)$ (from previous claim) $= \hat{\delta}(q_0, xz) \in F$ and so $xy^iz \in L$ $\square$

Finite Languages and Pumping Lemma

Question

Do finite languages really satisfy the condition in the pumping lemma?

Recall Pumping Lemma: If L is regular then there is a number p (the pumping length) such that $\forall w \in L$ with $|w| \geq p$, $\exists x, y, z \in \Sigma^*$ such that $w = xyz$ and

1. $|y| > 0$
2. $|xy| \leq p$
3. $\forall i \geq 0.\ xy^i z \in L$

Answer

Yes, they do. Let p be larger than the longest string in the language. Then the condition "$\forall w \in L$ with $|w| \geq p, \ldots$" is vaccuously satisfied as there are no strings in the language longer than p!

Using the Pumping Lemma

L regular implies that L satisfies the condition in the pumping lemma. If L is not regular pumping lemma says nothing about L!

Pumping Lemma, in contrapositive

If L does not satisfy the pumping condition, then L not regular.

Pumping ConditionNegation of the Pumping Condition

$$\cancel{\exists}\exists\forall p. \quad \cancel{\forall}\forall\exists w \in L. \text{ with } |w| \geq p \quad \cancel{\exists}\exists\forall x, y, z \in \Sigma^*. \; w = xyz$$

$$\left.\begin{array}{ll} (1) & |y| > 0 \\ (2) & |xy| \leq p \\ (3) & \forall i \geq 0. \; xy^iz \in L \end{array}\right\} \text{not all of them hold}$$

Equivalent to showing that if (1), (2) then (3) does not. In other words, we can find i such that $xy^iz \notin L$

Game View

Think of using the Pumping Lemma as a game between you and an opponent.

L	Task: To show that L is not regular
$\forall p.$	Opponent picks p
$\exists w.$	Pick w that is of length at least p
$\forall x, y, z$	Opponent divides w into x, y, and z such that $\lvert y \rvert > 0$, and $\lvert xy \rvert \leq p$
$\exists k.$	You pick k and win if $xy^k z \notin L$

Pumping Lemma: If L is regular, opponent has a winning strategy (no matter what you do).
Contrapositive: If you can beat the opponent, L not regular.
Your strategy should work for any p and any subdivision that the opponent may come up with.

Example I

Proposition

$L_{0n1n} = \{0^n 1^n \mid n \geq 0\}$ is not regular.

Proof.

Suppose L_{0n1n} is regular. Let p be the pumping length for L_{0n1n}.

- Consider $w = 0^p 1^p$
- Since $|w| > p$, there are x, y, z such that $w = xyz$, $|xy| \leq p$, $|y| > 0$, and $xy^i z \in L_{0n1n}$, for all i.
- Since $|xy| \leq p$, $x = 0^r$, $y = 0^s$ and $z = 0^t 1^p$. Further, as $|y| > 0$, we have $s > 0$.

$$xy^0 z = 0^r \epsilon 0^t 1^p = 0^{r+t} 1^p$$

Since $r + t < p$, $xy^0 z \notin L_{0n1n}$. Contradiction! □

Example II

Proposition

$L_{\mathrm{eq}} = \{w \in \{0,1\}^* \mid w$ *has an equal number of 0s and 1s*$\}$ *is not regular.*

Proof.

Suppose L_{eq} is regular. Let p be the pumping length for L_{eq}.

- Consider $w = 0^p 1^p$
- Since $|w| > p$, there are x, y, z such that $w = xyz$, $|xy| \leq p$, $|y| > 0$, and $xy^i z \in L_{\mathrm{eq}}$, for all i.
- Since $|xy| \leq p$, $x = 0^r$, $y = 0^s$ and $z = 0^t 1^p$. Further, as $|y| > 0$, we have $s > 0$.

$$xy^0 z = 0^r \epsilon 0^t 1^p = 0^{r+t} 1^p$$

Since $r + t < p$, $xy^0 z \notin L_{\mathrm{eq}}$. Contradiction! □

A Tale of two Proofs

Non Pumping Lemma

Suppose L_{eq} is recognized by DFA M with p states. Consider the input 0^p1^p. There exist j, k and state q such that

- $j < k$ and $\hat{\delta}(q_0, 0^j) = \hat{\delta}(q_0, 0^k) = q$
- Since $0^p1^p \in L_{\text{eq}}$, $0^k0^{(p-k)}1^p$ is accepted by M and so is $0^j0^{(p-k)}1^p$.
- But $0^j0^{(p-k)}1^p \notin L_{\text{eq}}$.

Pumping Lemma

Suppose L_{eq} is regular. Let p be pumping length for L_{eq}. Consider $w = 0^p1^p$. There exist x, y, z such that

- $w = xyz$, $|xy| \leq p$, $|y| > 0$: so for some r, s, t, $x = 0^r$, $y = 0^s$ and $z = 0^t1^p$, with $s > 0$.
- $xy^iz \in L_{\text{eq}}$ for all i: so $xy^0z \in L_{\text{eq}}$.
- But $xy^0z = 0^{p-s}1^p \notin L_{\text{eq}}$

Example III

Proposition

$L_p = \{0^i \mid i \text{ prime}\}$ is not regular

Proof.

Suppose L_p is regular. Let p be the pumping length for L_p.

- Consider $w = 0^m$, where $m \geq p + 2$ and m is prime.
- Since $|w| > p$, there are x, y, z such that $w = xyz$, $|xy| \leq p$, $|y| > 0$, and $xy^iz \in L_p$, for all i.
- Thus, $x = 0^r$, $y = 0^s$ and $z = 0^t$. Further, as $|y| > 0$, we have $s > 0$. $xy^{r+t}z = 0^r(0^s)^{(r+t)}0^t = 0^{r+s(r+t)+t}$. Now $r + s(r + t) + t = (r + t)(s + 1)$. Further $m = r + s + t \geq p + 2$ and $s > 0$ mean that $t \geq 2$ and $s + 1 \geq 2$. Thus, $xy^{r+t}z \notin L_p$. Contradiction! □

Example IV

Question

Is $L_{xx} = \{xx \mid x \in \{0,1\}^*\}$ is regular?

Suppose L_{xx} is regular, and let p be the pumping length of L_{xx}.

- Consider $w = 0^p 0^p \in L$.
- Can we find substrings x, y, z satisfying the conditions in the pumping lemma? Yes! Consider $x = \epsilon, y = 00, z = 0^{2p-2}$.
- Does this mean L_{xx} satisfies the pumping lemma? Does it mean it is regular?
  - No! We have chosen a bad w. To prove that the pumping lemma is violated, we only need to exhibit some w that cannot be pumped.
- Another bad choice $(01)^p(01)^p$.

Example IV

Reloaded

Proposition

$L_{xx} = \{xx \mid x \in \{0,1\}^*\}$ *is not regular.*

Proof.

Suppose L_{xx} is regular. Let p be the pumping length for L_{xx}.

- Consider $w = 0^p10^p1$.
- Since $|w| > p$, there are x, y, z such that $w = xyz$, $|xy| \leq p$, $|y| > 0$, and $xy^iz \in L_p$, for all i.
- Since $|xy| \leq p$, $x = 0^r$, $y = 0^s$ and $z = 0^t10^p1$. Further, as $|y| > 0$, we have $s > 0$.

$$xy^0z = 0^r\epsilon 0^t10^p1 = 0^{r+t}10^p1$$

Since $r + t < p$, $xy^0z \notin L_{xx}$. Contradiction! □

Lessons on Expressivity

Limits of Finite Memory

Finite automata cannot

- "keep track of counts": e.g., L_{0n1n} not regular.
- "compare far apart pieces" of the input: e.g. L_{xx} not regular.
- do "computations that require it to look at global properties" of the input. e.g. L_{prime} not regular.

...and pumping lemma provides one way to find out some of these limitations.