BU CS 332 – Theory of Computation

https://forms.gle/2d2ANba88rCLwZBC6

Lecture 7:

- Distinguishing sets
- Non-regular languages

Reading:

"Myhill-Nerode" note

Mark Bun

September 23, 2021

Motivating Questions

- How can we tell if we've found the smallest DFA recognizing a language?

  Last time: Introduced distinguishing set method

- Are all languages regular? How can we prove that a language is not regular?

A General Technique

$A = \{ w \in \{0,1\}^* \mid w \text{ ends with } 01 \}$

Definition: Strings x and y are **distinguishable** by L if there exists a "distinguishing extension" z such that exactly one of xz or yz is in L.

Either: $xz \in L$ and $yz \notin L$
or
$xz \notin L$ and $yz \in L$

Ex. $x = \varepsilon, \quad y = 0$

$z = 1 \qquad xz = \varepsilon 1 = 1 \notin A \qquad yz = 01 \in A$

Definition: A set of strings S is **pairwise distinguishable** by L if every pair of distinct strings $x, y \in S$ is distinguishable by L.

Ex. $S = \{\varepsilon, 0, 01\}$

$x = \epsilon,\ y = 0:\ z = 1 \qquad x = \epsilon,\ y = 01:\ z = \epsilon \qquad x = 0,\ y = 01:\ z = \epsilon$

Theorem: If S is pairwise distinguishable by L, then every DFA recognizing L needs at least $|S|$ states

Any DFA recog. A needs ≥ 3 states

A General Technique

Theorem: If S is pairwise distinguishable by L, then every DFA recognizing L needs at least $|S|$ states

$S = \{x_1, x_2, \ldots, x_k\}$ $\quad (k = |S|)$

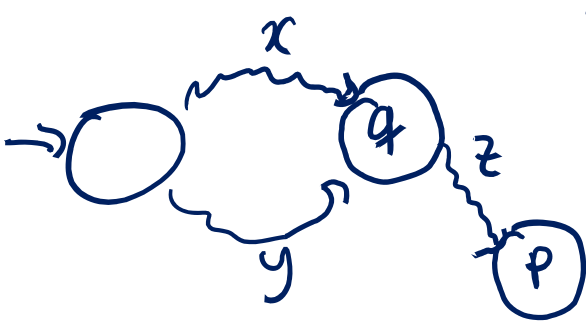
pigeons

Holes: DFA states $q_1, q_2, \ldots, q_r$ $\quad r < k$

Pigeon x_i is assigned to hole q_j if M ends in state q_j when reading x_i

Goal: Prove M does not recog. L

Proof: Let M be a DFA with $< |S|$ states. By the pigeonhole principle, there are $x, y \in S$ such that M ends up in same state on x and y



Since S is pairwise dist., $\exists z$ s.t. (wlog) $xz \in L$ and $yz \notin L$

$\Downarrow$ p needs to be an accept state

$\Downarrow$ p needs to be a reject state

p can't be both, so M can't recog. L.

Another Example

$B = \{ w \in \{0,1\}^* \mid |w| = 2 \}$

Theorem: If S is pairwise distinguishable by L, then every DFA recognizing L needs at least $|S|$ states

$S = \{ \epsilon, 0, 00, 000 \}$

Any DFA for B must tell apart:

Claim: Every DFA for B needs ≥ 4 states

Proof: Show S is pairwise dist. by B

$x = \epsilon, y = 0 : z = 0$

$x = \epsilon, y = 00 : z = \epsilon$

$x = \epsilon, y = 000 : z = 00$

1) $|w| > 2$

2) $|w| = 2$

3) $|w| = 1$

4) $|w| = 0$

Distinguishing Extension

Which of the following is a distinguishing extension for $x = 0$ and $y = 00$ for language $B = \{ w \in \{0,1\}^* \mid |w| = 2 \}$?

a) $z = \varepsilon$ ✓

b) $z = 0$

c) $z = 1$

d) $z = 00$

$0\varepsilon \notin B$, $00\varepsilon \in B$

$00 \in B$, $000 \notin B$

$01 \in B$, $001 \notin B$

$000 \notin B$, $0000 \notin B$

Historical Note

Converse to the distinguishing set method:

If L has no distinguishing set of size $> k$, then L is recognized by a DFA with k states

Myhill-Nerode Theorem (1958): L is recognized by a DFA with $\leq k$ states iff L does not have a distinguishing set of size $> k$

Non-Regularity

Theorem: If S is pairwise distinguishable by L, then every DFA recognizing L needs at least $|S|$ states

Corollary: If S is an **infinite** set that is pairwise distinguishable by L, then no DFA recognizes L

Contrapositive: $\exists$ a DFA M recog. L, then L has no infinite pairwise dist. set.

Proof. Let k = # states of M

By Thm, any pairwise dist. set S for L must have size $|S| \le k$, so no infinite dist. set exists.


LANGUAGE
DFA
INFINITE
DISTINGUISHING
SET
DFA
LANGUAGE
imgflip.com

h/t Islam

The Classic Example

Theorem: $A = \{0^n 1^n \mid n \geq 0\}$ is not regular

Proof: We construct an infinite pairwise distinguishable set

$S = \{\epsilon, 0, 00, 000, \ldots\} = \{0^n \mid n \geq 0\}$

Now show S is pairwise dist.:

Let $x, y \in S$ be arbitrary. Say $x = 0^n$, $y = 0^m$ $m, n \geq 0$, $m \neq n$

Distinguishing extension: $z = 1^n$

$xz = 0^n 1^n \in A$

$yz = 0^m 1^n \notin A$

Palindromes

$101 \in L \qquad 101^R = 101$

$1011 \notin L \qquad 1011^R = 1101 \neq 1011$

Theorem: $L = \{w \in \{0,1\}^* \mid w = w^R\}$ is not regular

Proof: We construct an infinite pairwise distinguishable set

$S = \{1^n 0 1^{n-1} \mid n \geq 1\}$

$S' = \{1 x 0 \mid x \in \{0,1\}^*\}$

$S'' = \{1^n 0 \mid n \geq 0\}$

S'' is pairwise dist.

Let $x = 1^n 0$, $y = 1^m 0$ $\quad m \neq n$

Let $z = 1^n$. Then $1^n 0 1^n \in L$

$1^m 0 1^n \notin L$.

S is pairwise dist: Let $x = 1^n 0 1^{n-1}$, $y = 1^m 0 1^{m-1}$, $n \neq m$

$z = 1 \qquad xz = 1^n 0 1^n \qquad y = 1^m 0 1^m$ does not work

$z = 1^{n-1} 0 1^n \qquad x = 1^n 0 1^{2n-2} 0 1^n \in L$

$y = 1^m 0 1^{m+n-2} 0 1^n \notin L$ ✓

Now you try!

Use the distinguishing set method to show that the following languages are not regular

$$L_1 = \{0^i 1^j \mid i > j \geq 0\}$$

$S = \{ 0^n \mid n \geq 0 \}$

Let $x, y \in S$ be arbitrary. Let $x = 0^n$, $y = 0^m$.
Assume wlog $n > m$.

Let $z = 1^m$. Then $xz = 0^n 1^m$ $(n > m)$ $\in L_1$
$yz = 0^m 1^m \notin L_1$

Now you try!

Use the distinguishing set method to show that the following languages are not regular

$$L_2 = \{ww \mid w \in \{0,1\}^*\}$$

$x = 11$

$y = 1111$

$yx = 111111$

$S = \{w \mid w \in \{0,1\}^*\}$

Let $x, y \in S$ be arb.

Let $z = x$. Then $xx \in L_2$

yx

Doesn't work

$S' = \{01^n \mid n \geq 0\}$

$x = 01^n$, $y = 01^m$ $m \neq n$

$z = 01^n$

$xz = 01^n01^n \in L_2$

$yz = 01^m01^n \notin L_2$

Reusing a Proof

Finding a distinguishing set can take some work…

Let's try to reuse that work!

How might we show that

$BALANCED$ = $\{w \mid w$ has an equal # of 0s and 1s$\}$

is not regular?

$\{0^n1^n \mid n \geq 0\}$ = $BALANCED \cap \{w \mid$ all 0s in w appear before all 1s$\}$

not regular   mystery   regular ✓

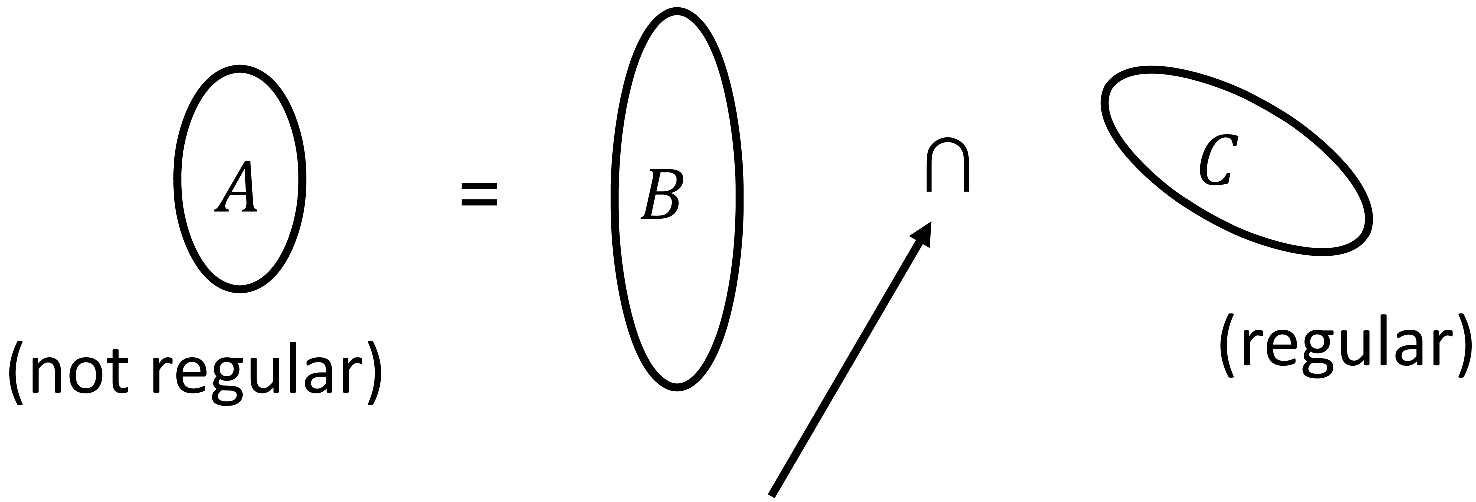
Assume (for contradiction) that BALANCED is regular

$\Rightarrow$ RHS is also regular (closure under $\cap$)

$\Rightarrow$ $\{0^n1^n \mid n \geq 0\}$ also regular $\times$.

Using Closure Properties

If A is not regular, we can show a related language B is not regular



any of $\{\circ, \cup, \cap\}$ or, for one language, $\{\neg, {}^{R}, {}^{*}\}$

<u>By contradiction:</u> If B is regular, then $B \cap C$ $(= A)$ is regular.

But A is not regular so neither is B!