BU CS 332 – Theory of Computation

Lecture 7:

- Distinguishing sets, non-regularity

Reading:

"Myhill-Nerode" note

Sipser Ch 1.4 (optional)

Mark Bun

February 16, 2021

Distinguishing Sets

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

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

distinguishing extension

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

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

$z = 1$

(either $xz \in L$ and $yz \notin L$
or
$xz \notin L$ and $yz \in L$)

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

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

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

$\exists$ S with $|S| = 3$ pairwise dist. by A $\Rightarrow$ every DFA for A requires ≥ 3 states

Another Example

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

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

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

1) $|w| > 2$: reject

2) $|w| = 2$: accept

3) $|w| = 1$: wait for 1 more char

4) $|w| = 0$: wait for 2 more chars

$S' = \{ \epsilon, 0, 11, 010 \}$

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) $z = 00$

$$\left.\begin{array}{ll} z = \varepsilon \Rightarrow & xz = 0 \notin B \\ & yz = 00 \in B \\ z = 0 \Rightarrow & xz = 00 \in B \\ & yz = 000 \notin B \\ z = 1 \Rightarrow & xz = 01 \in B \\ & yz = 001 \notin B \end{array}\right\} \text{all dist extensions}$$

$$\begin{array}{ll} z = 00 \Rightarrow & xz = 000 \notin B \\ & yz = 0000 \notin B \end{array}$$

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

Use this on HW

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

By contradiction: If L recognized by DFA of size k

then L does not have a pairwise dist. set of size $> k$

contradicts existence of infinite pairwise dist. set. $\times$

The Classic Example

$S = \{0^{2n}1^n \mid n \geq 0\}$

$x = 0^{2n}1^n$ $\quad z = 1^n$ $\quad xz = 0^{2n}1^{2n} \in L$

$y = 0^{2m}1^m$ $\quad yz = 0^{2m}1^{m+n} \notin L$

$m \neq n$

Theorem: $A = \{0^n1^n \mid n \geq 0\}$ is not regular

Proof: Construct an infinite pairwise distinguishable set

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

$S = \{0^n1^n \mid n \geq 0\}$ probably doesn't work

$x = 0^n1^n$ $\quad y = 0^m1^m$

WTS: $\forall x, y \in S$, $x \neq y$, $\exists z$ exactly one of xz or yz $\in L$

Let $x = 0^n$, $y = 0^m$ $\quad m \neq n$, $m, n \geq 0$ (arbitrary)

Take $z = 1^n$ $\quad xz = 0^n1^n \in L$

$yz = 0^m1^n \notin L$

$z = 1^m$ also a dist. ext.

$z = 0^n1^n$

$xz = 0^n0^n1^n$

$= 0^{2n}1^n \notin L$

$yz = 0^m0^n1^n$

$\in L$ if $m = 0$

$\notin L$ o.w.

Palindromes

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

Proof: Construct an infinite pairwise distinguishable set

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

$x = 0^n 1^n \qquad y = 0^m 1^m \qquad m \neq n$ Friend's move

$z = 0^n \qquad xz = 0^n 1^n 0^n \in L$ My move

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

$(yz)^R = 0^n 1^m 0^m \neq 0^m 1^m 0^n$

Now you try!

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

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

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

[$S' = \{0^n \mid n > 0\}$ also works]

let $x = 0^n$, $y = 0^m$ $n > m \geq 0$

$z = 1^m \Rightarrow xz = 0^n 0^m \in L$, $yz = 0^m 1^m \notin L$

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

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

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

$z = 0^n 1^n \Rightarrow xz = 0^n 1^n 0^n 1^n \in L$

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

$L_3 = \{1^{n^2} \mid n \geq 0\}$

$S = \{1^{n^2} \mid n \geq 0\}$

Let $x = 1^{n^2}$, $y = 1^{m^2}$, let $n > m \geq 0$

Let $z = 1^{2m+1}$

$yz = 1^{m^2+2m+1} = 1^{(m+1)^2} \in L_3$

$xz = 1^{n^2+2m+1} \notin L_3$

$n^2 < n^2 + 2m + 1 < n^2 + 2n + 1 = (n+1)^2$

Reusing a Proof

Finding a distinguishing set can take some work…

Let's try to reuse that work!

$0^n 1^n$ $\quad$ $011001 \in BALANCED$

How might we show that

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

is not regular?

(regular)

$L(0^* 1^*)$

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

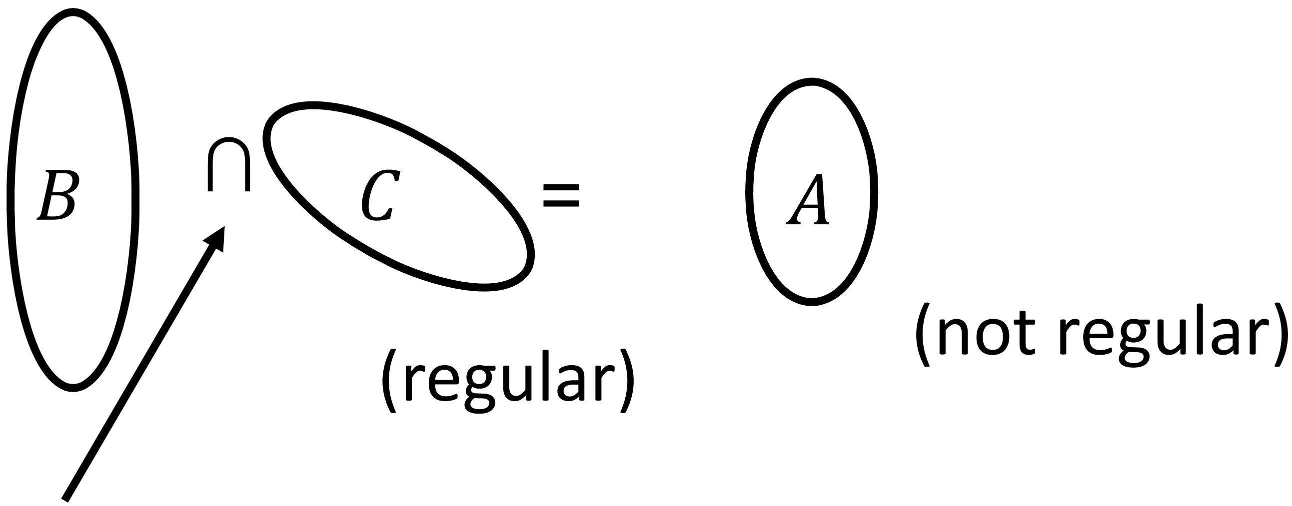
Assume for contradiction BALANCED is regular

$\Rightarrow$ RHS regular $\Rightarrow$ LHS regular ✗

(closure under $\cap$)

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!

Example

Prove $B = \{0^i 1^j | i \neq j\}$ is not regular using

- nonregular language

  $A = \{0^n 1^n | n \geq 0\}$ and

- regular language

  $C = \{w \mid \text{all 0s in } w \text{ appear before all 1s}\}$