BU CS 332 – Theory of Computation

https://forms.gle/amwrNaz3F7CiSsRa6

Lecture 7:

- Distinguishing sets
- Non-regular languages

Reading:

"Myhill-Nerode" note

Mark Bun

September 27, 2022

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.

$\in \Sigma^*$

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

A dist. extension $z = 1$

$xz = 1 \notin A$

$yz = 01 \in A$

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

$x = \epsilon \quad y = 0 \ : \ z = 1$ is a dist. ext.

$x = \epsilon \quad y = 01 \ : \ z = \epsilon$ is a dist. ext.

$(xz = \epsilon \notin A, \ yz = 01 \in A)$

$x = 0 \quad y = 01 \ : \ z = \epsilon$ is a dist. ext.

A General Technique

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

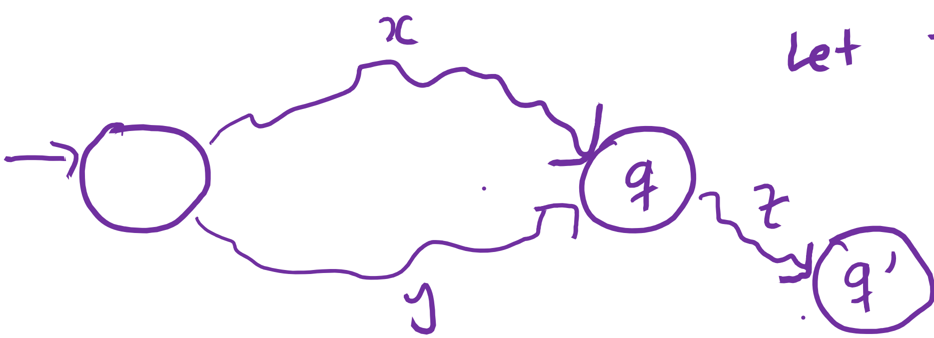
Proof: Let M be a DFA with $< |S|$ states. By the pigeonhole principle, there are $x, y \in S$ such that M ends up in same state on x and y

Pigeons: Elements $x_1, \dots, x_n$ of S $\quad n = |S|$

Holes: States of M

To assign pigeons to holes: Pigeon x goes to hole q if DFA M ends in state q after reading x.

By PHP: $\exists$ q s.t. two distinct x, y lead M to q



Let z be the dist. extension for x & y

Exactly one of xz or $yz \in L$

WLOG $xz \in L$, $yz \notin L$

Regardless of whether q' is an accept or non-accept state, M messes up on either xz or yz

$\Rightarrow$ M does not recognize L

Another Example

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

Want to show:
Every DFA recognizing B
needs at least 4 states

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

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

Claim: S is pairwise dist. by B

Proof.
$x = \epsilon \quad y = 0 : \quad z = 1$
$x = \epsilon \quad y = 00 : \quad z = \epsilon$
$x = \epsilon \quad y = 000 : \quad z = 11$
$x = 0 \quad y = 00 : \quad z = \epsilon$
$x = 0 \quad y = 000 : \quad z = 1$
$x = 00 \quad y = 000 : \quad z = \epsilon$

By Theorem, every DFA recognizing L needs at least $|S| = 4$ states.

Intuitively:
Any DFA for B needs to distinguish:
$|w| = 0$
$|w| = 1$
$|w| = 2$
$|w| = 3$

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) $z = \varepsilon$ $\quad 0\varepsilon \notin B \quad 00\varepsilon \in B$

✓ b) $z = 0$ $\quad 00 \in B \quad 000 \notin B$

✓ c) $z = 1$ $\quad 01 \in B \quad 001 \notin B$

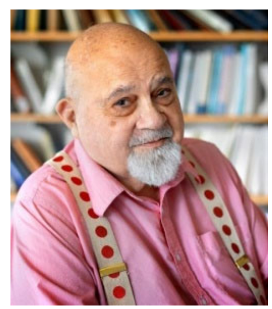
✗ d) $z = 00$ $\quad 000 \notin B \quad 0000 \notin B$

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 **if and only if** 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

Contrapositive: If $\exists$ a DFA for L using $< k$ states, then L has no pairwise dist. set of size $\geq k$

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

Contrapositive: If $\exists$ a DFA for L then L has no infinite pairwise dist. set

Proof of Corollary Contrapositive from Thm Contrapositive:

Suppose L is recognized by a DFA. Let k = # of states of this DFA. By Thm contrapositive, L has no pairwise dist. set of size $> k$. $\Rightarrow$ L has no infinite pairwise dist set.

The Classic Example

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

Proof: We construct an infinite pairwise distinguishable set

Idea: To recognize A, would need to distinguish between ϵ, 0, 00, 000, 0000, ...

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

Claim: S is an infinite pairwise dist. set by A

Let x and y be arbitrary distinct strings in S

$x = 0^n$, $y = 0^m$ $m \neq n$ $m, n \geq 0$

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

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

Palindromes

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

Proof: We construct an infinite pairwise distinguishable set

Let $S = \Sigma_1^*$ $x = 00$ $y = 001$???

$z = 100$ $xz = 00100$ $yz = 001100$

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

Claim: S is an infinite pairwise dist. set

Proof. Let $x = 0^n$, $y = 0^m$ $n \neq m$ arbitrary

Let $z = 10^n$

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

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

Now you try!

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

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

Your job: Construct an infinite set S that is pairwise dist. by L, i.e., $\forall\, x \neq y$ in S, $\exists$ a dist. extension z, i.e. exactly one of xz, yz in L

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

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

Set $z = 1^m$ $\quad$ Then $xz = 0^n 1^m \in L_1$

$yz = 0^m 1^m \notin L_1$

Now you try!

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

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

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

$S = \{0,1\}^*$

$x \neq y \in S$ $x = 00$ $y = 0000$

$z = x$ $z = 00$

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

First attempt: $z = 0^m$ $xz = 0^{2m} \in L_2$

$yz = 0^{m+n} \overset{?}{\in} L_2$

$z = 10^m1$ $xz = 0^m10^m1 \in L_2$

$yz = 0^n10^m1 \notin L_2$