Lecture 4: Closure Properties, Non-regularity and Myhill-Nerode Theorem

Date: August 31, 2023.

Homomorphism: A function $h : \Sigma^* \to \Gamma^*$ is a *homomorphism* if and only if $h(\epsilon) = \epsilon$, and for every $x, y \in \Sigma^*$, $h(xy) = h(x)h(y)$.

Proposition 1. *Consider homomorphisms $h_1, h_2 : \Sigma^* \to \Gamma^*$. $h_1 = h_2$ if and only if for all $a \in \Sigma$, $h_1(a) = h_2(a)$.*

Homomorphic and Inverse Homomorphic Images: Let $h : \Sigma^* \to \Gamma^*$ be a homomorphism, $A \subseteq \Sigma^*$, and $B \subseteq \Gamma^*$. Then,

$$h(A) = \{h(w) \mid w \in A\}$$
$$h^{-1}(B) = \{w \mid h(w) \in B\}$$

Theorem 2. *Regular languages are closed under both homomorphic and inverse homomorphic images.*

If $A \subseteq \Sigma^*$, $B \subseteq \Gamma^*$ are regular and $h: \Sigma^* \to \Gamma^*$ then $h(A)$ & $h^{-1}(B)$ are regular.

$h(A)$: Given w
Is $w \in h(A)$?
$\exists w \in A$ $h(w) = w$.

$\exists$ reg exp. r. $L(r) = A$.
$h(A)$: described by a reg exp where every symbol a in r is replaced by $h(a)$.

$$h(r) = \begin{cases} \emptyset & \text{if } r = \emptyset \\ \epsilon & \text{if } r = \epsilon \\ h(a) & \text{if } r = a \\ h(r_1)\, h(r_2) & \text{if } r = r_1 r_2 \\ h(r_1) + h(r_2) & r = r_1 + r_2 \end{cases} \qquad (h(r_1))^* \quad r = r_1^*$$

$h^{-1}(B)$: Given w.
Is $w \in h^{-1}(B)$?
$h(w) \in B$

$\exists M = (Q, \Gamma, \delta, s, F)$ s.t $L(M) = B$.
① Compute $h(w)$
② Check if $h(w) \in L(M)$

$N = (Q, \Sigma, \delta', s, F)$
$\delta'(q, a) = \hat{\delta}_M(q, h(a))$

$\hat{\delta}(q, \epsilon) = q$.
$\hat{\delta}(q, wa)$
$= \delta(\hat{\delta}(q,w), a)$

$\hat{\delta} : Q \times \Gamma^* \to Q$: $\hat{\delta}(q, w)$: state of M when it reads w from q.

Proposition 3. *For a language L, let* suffix$(L) = \{v \mid \exists u.\ uv \in L\}$. *If a language L is regular then* suffix(L) *is also regular.*

$L \subseteq \text{suffix}(L)$

$\bar{\Sigma} = \{\bar{a} \mid a \in \Sigma\}$

unbar : $(\Sigma \cup \bar{\Sigma})^* \to \Sigma^*$ unbar$(a) = a =$ unbar$(\bar{a})$
rembar : $(\Sigma \cup \bar{\Sigma})^* \to \Sigma^*$ rembar$(a) = a$
rembar$(\bar{a}) = \epsilon$

$$\text{suffix}(L) = \text{rembar}\big(\underbrace{\text{unbar}^{-1}(L) \cap \bar{\Sigma}^*\Sigma^*}_{\text{strings in } L \text{ where bars have put at the beginning}}\big)$$

Last k symbols: For a string $w \in \Sigma^*$, $\mathsf{last}_k(w)$ is the last k symbols in the string w. This can be formally defined as

$$\mathsf{last}_k(w) = \begin{cases} w & \text{if } |w| < k \\ v & \text{if } w = uv \text{ where } u \in \Sigma^* \text{ and } v \in \Sigma^k \end{cases}$$

Problem 1. Show that any DFA recognizing

$$L_k = \{w \in \{0,1\}^* \mid \mathsf{last}_k(w) = 1u \text{ where } u \in \{0,1\}^{k-1}\}$$

has at least 2^k states.

Given $w \in \{0,1\}^*$
Yes if w has 1 exactly k positions from end!

$M_k = (Q, \{0,1\}, \delta, s, F)$

$Q = \{u \mid |u| \le k\}$. $|u| = k$

$F = \{1u \mid |u| = k-1\}$.

$s = \epsilon$. | 0^k

$\delta(u, a) = \mathrm{last}_k(ua)$

Observation: $\forall w,\ w \in L_k$ iff $0^k w \in L_k$.

$\hat{\delta}(u, w) = \mathrm{last}_k(uw)$

A: $\exists$ DFA M $L(M) = A$ and $\hat{\delta}(s,u) = \hat{\delta}(s,v)$

$\Rightarrow \forall w\ [\hat{\delta}(s, uw) = \hat{\delta}(s, vw)]$ $\{uw, vw\} \subseteq A$ or $\{uw, vw\} \cap A = \emptyset$

Contrapositive: A: $\boxed{\exists w\ |\{uw, vw\} \cap A| = 1}$ $u \not\equiv_A v$

$\Rightarrow \forall$ DFA M. $L(M) \neq A$ or $\hat{\delta}(s,u) \neq \hat{\delta}(s,v)$

$\forall$ DFA M. $L(M) = A \Rightarrow \hat{\delta}(s,u) \neq \hat{\delta}(s,v)$

$F = \{u \mid |u| = k\}$, $|F| = 2^k$.

Claim: F is a fooling set for L_k.

Proof: Consider $u, v \in F$. $u \neq v$.

$\exists$ position i, $u[i] \neq v[i]$ WLOG. $u[i] = 1$, $v[i] = 0$.

Take $w = 0^{i-1}$.

$\mathrm{last}_k(uw) = 1w'0^{i-1}$ $\qquad \mathrm{last}_k(vw) = 0\,v[i+1, k+1]\,0^{i-1}$

$w' \to u[i+1, k+1]$

$uw \in L_k$, $vw \notin L_k$.

Language Congruence: For any language $L \subseteq \Sigma^*$, let $\equiv_L \subseteq \Sigma^* \times \Sigma^*$ be the relation defined as

$$x \equiv_L y \text{ iff } \forall z \in \Sigma^*.\ xz \in L \leftrightarrow yz \in L$$

Fooling Set for a language $L \subseteq \Sigma^*$ is a set $F \subseteq \Sigma^*$ such that for every $x, y \in F$, if $x \neq y$ then $x \not\equiv_L y$. Or, for every $x, y \in F$, if $x \neq y$ then there is a z such that $|\{xz, yz\} \cap L| = 1$.

Proposition 4. *If F is a fooling set for L then any DFA recognizing L must have at least $|F|$ states.*

Problem 2. Prove that $L_{0^n1^n} = \{0^n1^n \mid n \geq 0\}$ is not regular.

$B = \{ w \in \{0,1\}^* \mid$ # 01 substrings w = # 10 substrings $w \}$.

010 – one 01 substring = one 10 substring $\in B$.

010101 . . .

Find an infinite sized fooling set for L_{0n1n}.

$F = \{0^i \mid i \geq 0\} = \{0\}^* = L(0^*)$

Claim: F is a fooling set

Consider $u = 0^i$, $v = 0^j \in F$, $u \neq v$ $(i \neq j)$

Take $w = 1^i$. $uw = 0^i1^i \in L_{0n1n}$, $vw = 0^j1^i \notin L_{0n1n}$

Problem 3. Prove that the language $A = \{uv \mid |u| = |v| \text{ and } u \neq v\}$ is not regular.

If L is regular then every fooling set is finite.

If L has an infinite fooling set then L is not regular.

Converse: If L is not regular then L has an infinite Fooling set

If ~~A~~ the max-sized fooling set of L is finite then L is regular.

Proposition 5. *For any language L, $\equiv_L$ is a* congruence.

$\equiv_L$ — Equivalence relation

and $\quad x \equiv_L y \Rightarrow \forall z.\ xz \equiv_L yz$.

Equivalence Classes: For an equivalence relation $\equiv \subseteq A \times A$, the equivalence class of $a \in A$ is the set

$$[a]_\equiv = \{b \in A \mid a \equiv b\}.$$

1. Equivalence classes of $\equiv$ form a partition of A. That is, for any $a, b \in A$, either $[a]_\equiv = [b]_\equiv$ or $[a]_\equiv \cap [b]_\equiv = \emptyset$.
2. The **index** of equivalence relation $\equiv$ is the number of equivalence classes, i.e.,

$$\#(\equiv) = |\{[a]_\equiv \mid a \in A\}|.$$

Theorem 6 (Myhill-Nerode). *A language L is regular if and only if $\#(\equiv_L)$ is finite.*

If L is regular then $\#(\equiv_L)$ is finite
Proved.

If $\#(\equiv_L)$ is finite then L is regular.

DFA for L. $M_{\equiv_L}(Q, \Sigma, \delta, s, F)$

$Q = \{[w]_\equiv \mid w \in \Sigma^*\}$.

$s = [\epsilon]_\equiv$ $\qquad$ $u \equiv_L v \quad [u] = [v]$?

$\delta([u]_\equiv, a) = [ua]_\equiv$ $\qquad$ But $[ua] \not\equiv [va]$

$F = \{[w]_\equiv \mid w \in L\}$.

Max fooling set size $=$ # states in automaton $M_{\equiv_L}$.
Thus $M_{\equiv_L}$ is the DFA with fewest states that recognizes L.