CSE 396 — lecture Thu 2/28 — Spr 2019

Top Hat #1245

Helpful Notation: $\underline{L(x)}$ = '$x \in L$' as a Boolean function.
Then we can abbreviate $x \in L \iff w \in L$ as $L(x) = L(w)$

And $L(x) \neq L(w)$ abbreviates $\underline{x \in L \text{ XOR } w \in L}$. Now let us define:

compare text ch1 exercises around 45-50

$\underline{x \sim_L y}$ to mean $(\text{for all } z \in \Sigma^*)\ L(xz) = L(yz)$

$x \not\sim_L y$ hence means $(\exists z \in \Sigma^*)\ L(xz) \neq L(yz)$.

In the latter case, say x and y are "distinctive" for L.

Now call a set $S \subseteq \Sigma^*$ a {distinctive set / $\underline{\text{PD set}}$ / pairwise distinguishing set} for L

if for all $\underbrace{\text{distinct } x, y \in S}_{\text{means } x \neq y.}$, $\underbrace{x \not\sim_L y}_{\text{means something more.}}$.

Key Insight #3: Suppose S is a PD set for L of size k. Then any DFA M such that $L(M) = L$ must have at least k states. indeed, must process the members of S to different states.

Proof: Suppose we have M with $k-1$ or fewer states s.t. $L(M) = L$. Then, by the Pigeonhole Principle (PHP) there must be two different strings $x, y \in S$ such that $\delta^*(s, x) = \delta^*(s, y)$. I.e., such that M processes x and y to the same state.

But by S being PD for L, $x \not\sim_L y$, so there is a string z s.t. $L(xz) \neq L(yz)$. By previous "Insights," M must be wrong on xz or wrong on yz. Contradiction.

Myhill - Nerode Theorem [1958]: "($\Rightarrow$)" Suppose S is an infinite PD set for a language L. Then L is not regular.

Myhill: John, UB Math †1987. Nerode: Anil, still at Cornell!

"($\Leftarrow$)" [And conversely: if L is not regular then there is ALWAYS an infinite PD set for it.]

Proof of $\Rightarrow$: Suppose L were regular. Then there would be a DFA M s.t. $L(M) = L$. M would have some finite number k of states. But S has (more than) $k+1$ strings... ☒

MNT says: ~~If~~ L is ~~non-regular~~ a language such that

there exists an infinite $S \subseteq \Sigma^*$ such that] L has an infinite PD set

for all $x, y \in S$, $x \neq y$] S is PD for L

there exists a string z s.t. $L(xz) \neq L(yz)$] $x \not\equiv_L y$

then L is not regular.

How to make this into a "Script for Proofs."

Take $S =$ ______. "Clearly S is infinite." (if it is really clear)

Let any $x, y \in S$, $x \neq y$ be given. Then, (based on how we defined S) (wlog) we can helpfully write $x =$ ____ and $y =$ ____ where ______.

Take $z =$ ______. Then $L(xz) \neq L(yz)$ because ______

$\therefore x \not\equiv_L y$

Since $x, y \in S$ are arbitrary, $\therefore$ S is PD for L.

Thus S is PD for L, and since S is infinite, L is not regular by MNT.

Example: $L = \{a^n b^n : n \geq 0\}$ (k, k, $k \geq 0$). Prove via MNT that L is not regular.

Take $S = \underline{a^*}$. Clearly S is infinite. Let any $x, y \in S$, $x \neq y$, be given.

Then we can write $x = a^m$, $y = a^n$ where $m, n \geq 0$ and $m \neq n$.

Take $z = b^m$. Then $xz = a^m b^m \in L$ but $yz = a^n b^m \notin L$ since $m \neq n$.

Thus $L(xz) \neq L(yz)$ and since $x, y \in S$ are arbitrary, S is an infinite PD set for L.

Thus L is not regular by MNT.

x is a palindrome: abba yes, ε yes, bab yes; abbb no, ba no.

Example 2: $L = \{x \in \{a,b\}^* : x = x^R\}$. Prove that L is not regular.

Take $S = \underline{a^*b}$. Clearly S is infinite. Let any $x, y \in S$, $x \neq y$ be given. Then we can write $x = a^m b$, $y = a^n b$, where $m \neq n$. Take $z = \underline{a^m}$. Then $xz = a^m b a^m \in L$, but $yz = a^n b a^m$ which is not a palindrome because $m \neq n$. (...)

Example 3: $L = \{x \in \{a,b\}^* : \#a(x) \geq \#b(x)\}$.

Take $S = a^*$. Clearly S is infinite. Let any $x, y \in S$, $x \neq y$, be given. Then we can write $x = a^m$, $y = a^n$, where wlog $m < n$. ie. without loss of generality we can let "x" refer to the shorter string.

Take $z = b^n$. Then $xz = a^m b^n$ is not in L because m is not $\geq n$.

But $yz = a^n b^n$ is in L since $n \geq n$. Thus $L(xz) \neq L(yz)$, and we conclude L is not regular as before... $\boxtimes$

Added:

Recitations will cover a main way one can go astray with MNT proofs.

Define $L = \{x \cdot y : \#0(x) = \#1(y)\}$. [The alphabet is just $\{0,1\}$, dot is concatenation

Take $S = 0^*$, clearly infinite. Let any $x, y \in S$, $x \neq y$, be given. Then we can write $x = 0^m$, $y = 0^n$, where $m \neq n$. Indeed we can say that wlog. $m < n$. Take $z = 1^m$. Then $xz = 0^m 1^m \in L$ but $yz = 0^n 1^m \notin L$ since $n > m$.

Oops, no: this is a "poof" not a proof. First of all, let's rename the dummy variables x, y in the definition of L to avoid a "symbol clash" with x, y, z in MNT:

$$L = \{uv : u, v \in \{0,1\}^*, \#0(u) = \#1(v)\}.$$

To see that $0^n 1^m$ DOES belong to L, use $n > m$ to take $u = 0^m$ and $v = 0^{n-m} 1^m$. Then u has m 0s and v has m 1s, so $\#0(u) = \#1(v)$. If we had $n < m$ we would do $u = 0^n 1^{m-n}$, $v = 1^n$ instead. So the proof does NOT get $L(xz) \neq L(yz)$ and so it fails. Indeed the proof must fail: L = all of $\{0,1\}^*$, which is regular! If we wrote $L = \{w : w$ can be broken as $w =: uv$ such that $\#0(u) = \#1(v)\}$ we could see this better.