CSE 396 Lecture Thu Mar 2 Spr 2017

PS4 will be posted later today. Prelim II → Mar 14.

Given any language $L \subseteq \Sigma^*$, define the relation

$x \sim_L y$ if for all $z \in \Sigma^*$, $xz \in L \Leftrightarrow yz \in L$.

Then $\sim_L$ is an equivalence relation: For all x, y, z (and/or ^)

✓ • $x \sim_L x$ because trivially "$xz \in L \Leftrightarrow xz \in L$."

✓ • $x \sim_L y \Leftrightarrow y \sim_L x$ because x and y are interchangeable in the defn of $\sim_L$.

✓ • if $x \sim_L y \equiv$ for all $u \in \Sigma^*$: $xu \in L \Leftrightarrow yu \in L$
and $y \sim_L z \equiv$ for all $v \in \Sigma^*$: $yv \in L \Leftrightarrow zv \in L$
then $x \sim_L z$. $\Leftarrow$ for all $w \in \Sigma^*$: $xw \in L \Leftrightarrow zw \in L$ (Think $v = u$, $w = y$)

Note: $x \not\sim_L y$ means $[\exists z \in \Sigma^*]$ $xz \in L$ ⇔ (yucky symbol) $yz \in L$.

ie. $xz \in L$ XOR (still yucky) $yz \in L$

ie. $L(xz) \neq L(yz)$

Regarding $L \subseteq \Sigma^*$ as the same as its Boolean characteristic function "$L(x)$".

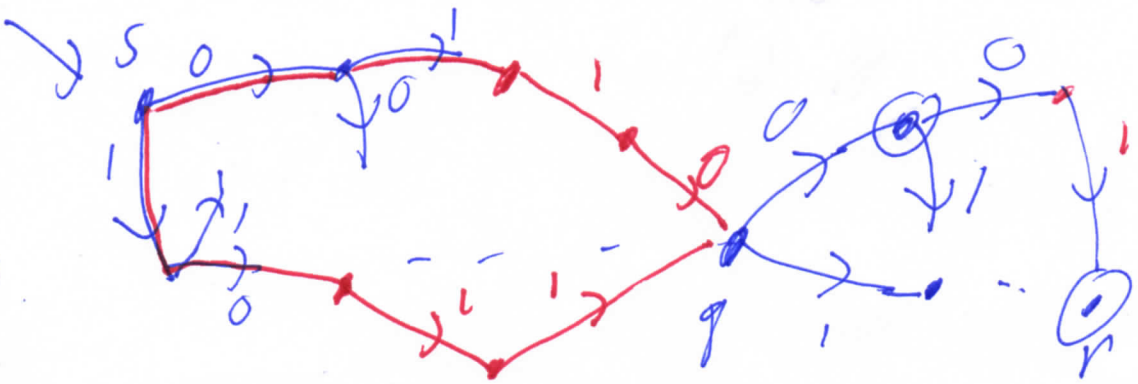
Defn: A subset $S \subseteq \Sigma^*$ is a distinctive / pairwise-distinct (PD) set for L if for all pairs $x, y \in S$, $x \neq y$, we have $x \not\sim_L y$. Not necessarily $S \subseteq L$!

Observation: If a DFA M processes two strings $x, y \in \Sigma^*$ to the same state q, then $x \sim_{L(M)} y$.

Picture

$x = 0110$

$y = 1011$



Whatever z comes next, xz and yz get processed to the same state r.
eg if $z = 001$, both are accepted, if $z = 1$, both are rejected.

Contrapositive: If $x \not\sim_{L(M)} y$ then M must process x and y to different states.

Key Point: If S is a PD set for $L(M)$ then M must process every string in S to a different state.
In particular, if $|S| = k$, M needs at least k states.

Proof: Suppose M has $k-1$ (or fewer) states. Since $|S| = k$, there exist two distinct strings $x, y \in S$ that M processes to the same state $q \in Q$. But then $x \sim_{L(M)} y$ from Observation which contradicts the def^n of PD saying $x \not\sim_{L(M)} y$. ☒

"Pigeonhole Principle"

Theorem (John Myhill, UB †1987, Anil Nerode)

If there is an infinite PD set S for a language L, then L is not regular.

(And vice-versa: if L is not regular, such S exists, and "L is regular" $\Leftrightarrow$ all PD sets for L are finite)

Proof: (of just the "If" part) Suppose L were regular. Then there would be a DFA $M = (Q, \Sigma, \delta, s, F)$ s.t. $L(M) = L$. Q would have some finite number K of states. But S, being PD, has $K+1$ mutually distinctive strings, so M would need at least $K+1$ states ✗ ⊗.

MNT, "unpacking" the defn's of PD, $x \not\equiv_L y$:

Given L, suppose we can find an infinite set S such that (For all $x, y \in S$, $x \neq y$)

"For all" ≡ PD.

$x \not\equiv_L y$ —

(there exists $z \in \Sigma^*$) s.t. $L(xz) \neq L(yz)$.

Then L is not regular.

"Proof Script": Take $S = 0^*$. Then S is infinite clearly

Let any $x, y \in S$ $(x \neq y)$ be given. Then based on the form of S, we can helpfully write

$x = 0^m$ and $y = 0^n$ for some numbers m, n ... where $m \neq n$.

without loss of generality. Take $z = 1^m$.

Then $L(xz) \neq L(yz)$ because $xz = 0^m 1^m \in L$, but $yz = 0^n 1^m \notin L$. (because $m \neq n$)

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

$L_1 = \{0^n 1^n : n \geq 0\}$.
$= \{0^m 1^n : m = n\}$.

$xz \in L$ but $yz \notin L$, so $L(xz) \neq L(yz)$. This is general, so L is not regular.

$L_2 = \{x \in \{0,1\}^* : \#0(x) > \#1(x)\}$: Take $S = 0^*$ (clearly infinite)

Let any $x, y \in S$, $x \neq y$, be given. Then there are numbers $m, n \in \mathbb{N}$ such that $x = 0^m$, $y = 0^n$, where without loss of generality $m < n$.

Take $z = 1^m$. Then $xz = 0^m 1^m \notin L_2$ because the number of 0s is not greater, while $yz = 0^n 1^m \in L_2$ since $n > m$.

So $L_2(xz) \neq L_2(yz)$, so L_2 is not regular by MNT. ☒
(since S is infinite and PD for L_2)