Learning a DFA

Sunday, September 29, 2019 8:25 PM

Seeing an unknown DFA in action, i.e. accepting some x's and not others, figure out the DFA.

How to formalize this?

Can ask: does $x \in L$?

Not enough!

For any finite list of x's there are multiple distinct DFAs.

How about: Find smallest DFA that accepts $\{x_1, x_2, \ldots\}$ — FINITE.

and does not accept $\{y_1, y_2, \ldots\}$

and does not accept $\{Y_1, Y_2, \ldots\}$

This is a computationally HARD problem. (NP-hard).

But doable: enumerate all DFA's in order of size, and check each one till you find one that works.

TIME: EXPONENTIAL in size of smallest DFA.

LEARNING MODEL.

- can ask $x \in L$? < YES / NO

- can also ask: is D the DFA?
  - YES ✓
  - NO – counterexample string.

Problem: Using only the above MEMBERSHIP and EQUIVALENCE queries, find the unknown DFA (or one that accepts the same language)

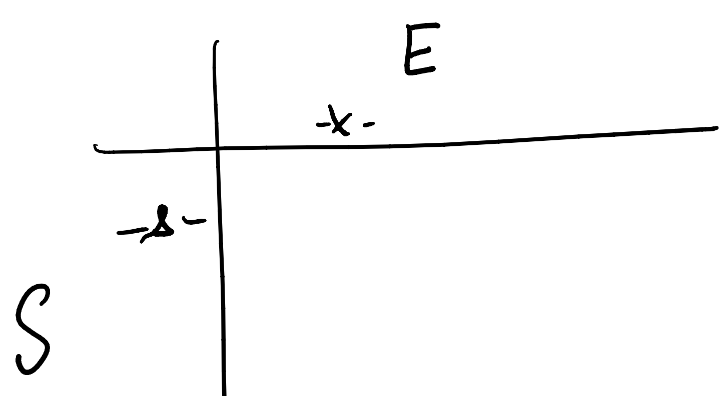
Sound Familiar?!

ANGLUIN'S ALGORITHM.

For simplicity assume $\Sigma = \{0, 1\}$.

Maintain set of candidate states and set of query strings

Observation Table.



S labeled by bit strings.

S is prefix closed

$1101 \in S$

$\rightarrow 110, 11, 1 \in S$

S

$1101 \ldots$

$\Rightarrow 110, 11, 1 \in S$

E is suffix closed

$1101 \in E$

$\Rightarrow 101, 01, 1 \in E$

$S \cdot \Sigma$

$$T(s, x) = \begin{cases} 1 & s \cdot x \in L \\ 0 & \text{o.w.} \end{cases}$$

Idea: distinct rows of S are states

closed: rows of $S \cdot \Sigma$ are included in S.

consistent: if $row(s_1) = row(s_2)$, then $\forall a \in \Sigma$
$row(s_1 \cdot a) = row(s_2 \cdot a)$.

GOAL: Find a closed, consistent observation table.

If not closed, move rows of $S \cdot A$ to S.

If not consistent, add to E.

Keep S prefix closed, E suffix closed.

Given closed, consistent table, create

Given closed, consistent table, create DFA from it:

$Q = \{ row(s) : s \in S \}$

$q_0 = row(\varepsilon)$

$F = \{ row(s) : T(s) = 1 \}$

$\delta(row(s), a) = row(s \cdot a)$.

Propose this DFA.
If counterexample, add to S, with all its prefixes.

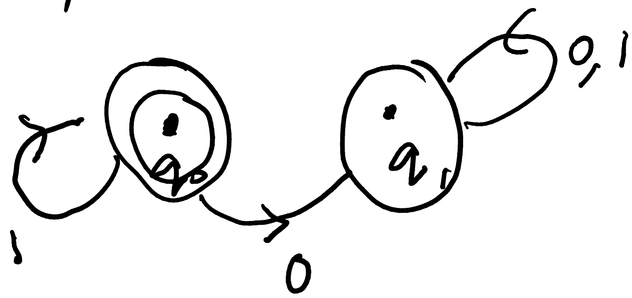
<u>Example</u>:

	ε
ε	1
0	0
1	1

not closed

	ε
ε	1
0	0
1	1
00	0
01	0

closed, consistent



counterexample: 000 $\quad T(000) = 1$

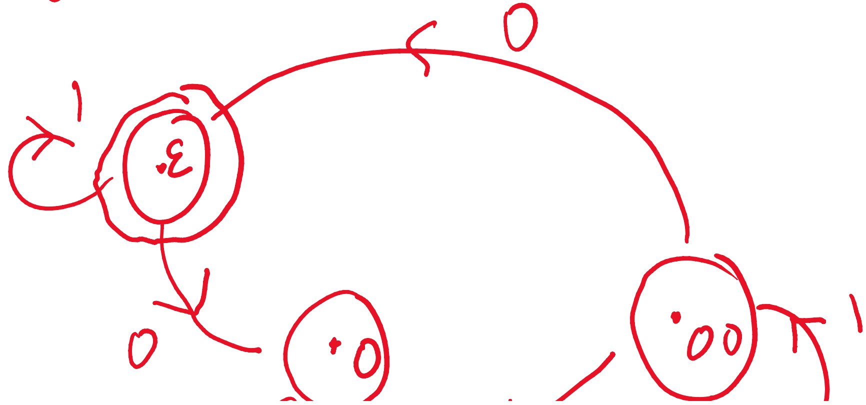
	ε	0
ε	1	0
0	0	0
00	0	1
000	1	0
1	1	0
01	0	0
001	0	1
0000	0	0
0001	1	0

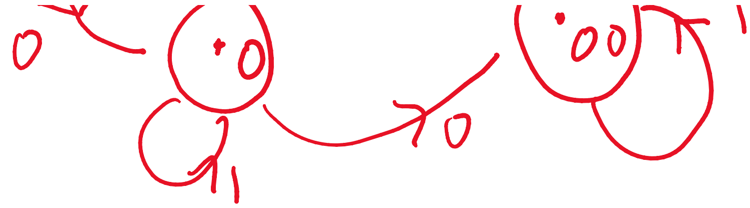
closed

not consistent since $row(0.0) \neq row(00.0)$

so add 0 to E

closed, consistent.





accepts iff #0's is divisible by 3.

Lemma 1 For a closed consistent table

$\forall s \in S \cup S\cdot\Sigma$, $\delta(q_0, s) = row(s)$

Lemma 2. For a closed, consistent table

$\forall s \in S \cup S\cdot\Sigma$, $e \in E$, $\delta(q_0, s\cdot e) \in F$ iff $T(s\cdot e) = 1$

Lemma 3. Any DFA consistent with T must have at least $|\{row(s) : s \in S\}|$) states.