Comp 330 - Tutorial 6 - Recorded

Learning automata

Q: How do I design a <u>learning algorithm</u> which by observing string membership can automatically generate an automaton?

Learning algo.

$$\mathcal{O} \rightarrow \boxed{A} \rightarrow \text{Automaton } M \text{ which "fits" } \mathcal{O}$$

i.e. $L(M) = L$

$01 \in L$
$010 \notin L$
...

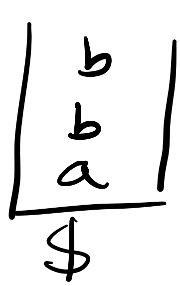
Recall that in the automata theory portion of 330, the "classic" question/exercise is something like:

Given $L \subseteq \Sigma^*$, where L is described using set builder notation i.e. using a predicate function where $P(w) \iff w \in L$, we would ask to design an automaton M which accepted L.

Ex $L_1 = \{ w \in \{a,b\}^* : w \text{ is a } \underbrace{\text{palindrome}}_{P(w)} \}$

The general strategy to design a PDA M s.t. $L(M) = L_1$ would be to *implement*, using the PDA architecture, the logic of $P(w)$.

abb, bb a
aba



$\rightarrow$ Leverage this intuition to design my PDA.

Then I conclude that L_1 is CF, & that the logic of P is implementable using a PDA.

What if L is not described using $P(w)$ (or maybe P is too complicated), but rather we are given string membership to L.

$\mathcal{O}$ Observation Table

Unknown lang (i.e. no $P(w)$)

String	Membership to L
ε	1
a	0
b	1
aa	0
$\vdots$	$\vdots$

$T(w) = 1 \iff w \in L$ according to $\mathcal{O}$

$\mathcal{O} \to \mathcal{A} \to M$ s.t. $L(M) = \{ w : \underset{1}{T(w) \overset{\shortparallel}{}} \}$

learning algorithm

We will study a learning algorithm $\mathcal{A}$ (from late 80s) where we assume L is unknown (i.e. no $P(w)$ / $P(w)$ is too complicated) AND <u>regular</u>. $\therefore$ The automata which $\mathcal{A}$ will produce will be minimal DFA.

TODO

minimal $\to$ fix this to just say DFA

Active learning algorithm

Given some unknown regular language L, we assume that we are interfacing with a teacher T which can give us feedback on the output of A.

$$\emptyset \to A \to M', \text{ candidate DFA}$$

① $M' \to T$
- $\nearrow$ yes, $L(M') = L$ & M' is min
- $\searrow$ No, M' is not correct + counterexample

$$w \in \big(\underbrace{L - L(M')}_{\text{false negative}} \cup \underbrace{L(M') - L}_{\text{false positive}}\big)$$

② $w \to T$
- $\nearrow$ y, $w \in L$
- $\searrow$ N, $w \notin L$.

Why not just use T?

<u>Ex</u> You're a ML engineer & want to design a classifier (DFA) which checks whether a research article idea is interesting / proposes novel research idea. Imagine you're designing an <u>expert system</u> => program w/ a collection of if-then-else statements. You go to an expert researcher: 1. Really good at classifying interesting papers 2. Has some intuition on what makes a paper interesting. How do you design the system based on expert researcher (T)?

<u>Description of the L^* alg. for active learning of regular languages</u>

Observation table $\mathcal{O}$, 2D array

$\Sigma \neq \emptyset$, $\underset{\text{unknown}}{L} \subseteq \Sigma^*$, $\mathcal{O}$: $S \subseteq \Sigma^* \to$ rows
$E \subseteq \Sigma^* \to$ colums

$$\mathcal{O} : \overbrace{(S \cup S \cdot \Sigma)}^{\text{rows}} \times \overbrace{E}^{\text{cols}} \to \{0, 1\}$$

where $\mathcal{O}(s, e) = T(s \cdot e)$

$s \in S \cup S \cdot \Sigma \quad e \in E$

teacher $T(w) = 1$ $\Leftrightarrow$ $w \in L$

Ex $S = \{ \varepsilon, b \}$ $\quad E = \{ \varepsilon, a, aa \}$

$\underline{Ex}$ $S = \{\varepsilon, b\}$ $E = \{\varepsilon, a, aa\}$

rows of $\mathcal{O}$ $= \{\varepsilon, b\} \cup \{\varepsilon, b\} \cdot \{a, b\}$

$= \{\varepsilon, b, a, ba, bb\}$

cols of $\mathcal{O}$ $= \{\varepsilon, a, aa\}$

$\mathcal{O}$

		ε	a	aa
S	ε			
	b			$T(b \cdot aa) = 1 \iff baa \in L$
$\cup$ $S \cdot \Sigma$	a			
	ba			
	bb	$T(bb \cdot \varepsilon) = 0 \iff bb \notin L$		

Example of $\mathcal{O}$

		$\overbrace{\varepsilon}^{E}$	$\overbrace{a}^{E}$	
S	ε	0	0	$row(\varepsilon) = 00$
	a	0	1	
	aa	1	0	$row(aa) = 10$
$S \cdot \Sigma$	b	0	$\boxed{0}$	$\rightarrow T(b \cdot a) = 0 \iff ba \notin L$
	ab	0	0	
	aaa	0	0	
	aab	0	0	$row(aab) = 00$

How do I construct a DFA from $\mathcal{O}^*$ (*) $\Rightarrow$ Assuming $\mathcal{O}$ satisfies some "nice" properties which allow us to create a corresponding DFA.

States of $M_{\mathcal{O}} = (Q, \Sigma, \delta, q_0, F)$

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

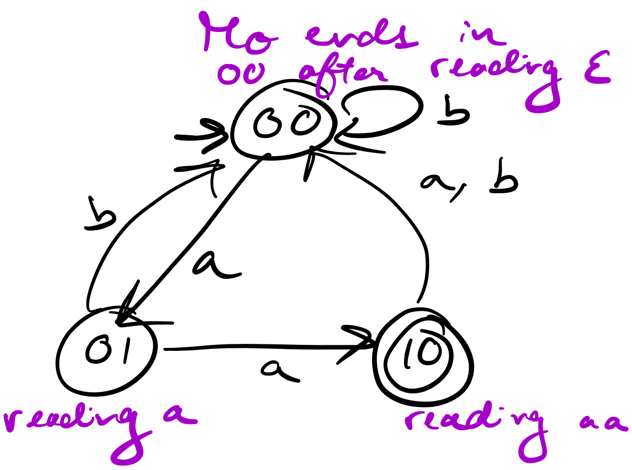
$q_0 := \text{row}(\varepsilon)$

$F := \{ \text{row}(s) : s \in S, \; T(s\cdot\varepsilon) = \mathcal{O}(s,\varepsilon) = 1 \}$

$\delta(\underbrace{\text{row}(s)}_{\in Q}, \underbrace{\sigma}_{\in \Sigma})$

		E: ε	a
S	ε	0	0
	a	0	1
	aa	1	0
$S\cdot\Sigma$	b	0	0
	ab	0	0
	aaa	0	0
	aab	0	0

$T(s\cdot\varepsilon) = T(s) = 1 \Leftrightarrow s \in L$



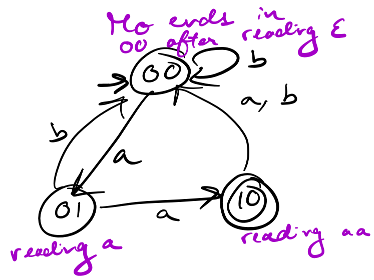
DFA $M_{\mathcal{O}}$ induced by $\mathcal{O}^*$

Special properties of $\mathcal{O}$:

	ε	a
ε	1 0	0
a	0	1
aa	1	0
b	0	0
ab	0	0
aaa	0	0
aab	0	0
bb	0	0
ba	1	0

S

$S \cdot \Sigma$

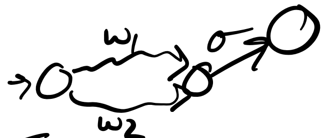


→ $\mathcal{O}$ is not closed

→ $\mathcal{O}$ is not consistent

Def $\mathcal{O}$ is closed if [Prevent undefined transitions]

$$\forall\, t \in S \cdot \Sigma \;\; \exists\, s \in S \text{ s.t. } row(t) = row(s)$$



Def $\mathcal{O}$ is consistent if [Ensures that M_O "behaves" like a DFA]

$$\forall\, s_1, s_2 \in S \text{ if } row(s_1) = row(s_2) \Rightarrow \forall\, \sigma \in \Sigma \;\; row(s_1 \cdot \sigma) = row(s_2 \cdot \sigma)$$

An example run of the L^* alg

0. Initial $\mathcal{O}_0$, $S = E = \{\varepsilon\}$

$\Sigma = \{a, b\}$

		$\overbrace{\varepsilon}^{E}$
S	ε	1
$S\cdot\Sigma$	a	0
	b	0

Is $\mathcal{O}_0$ closed? Is $\mathcal{O}_0$ consistent?

$\mathcal{O}_0$ is not closed, b/c $row(a) = 0 \notin \{row(s) : s \in \underset{\{\varepsilon\}}{\underset{\Downarrow}{S}}\}$

When $\mathcal{O}_0$ is not closed, $S \leftarrow S \cup \{a\}$

1. $\mathcal{O}_1$ where $E = \{\varepsilon\}$ $S = \{\varepsilon, a\}$

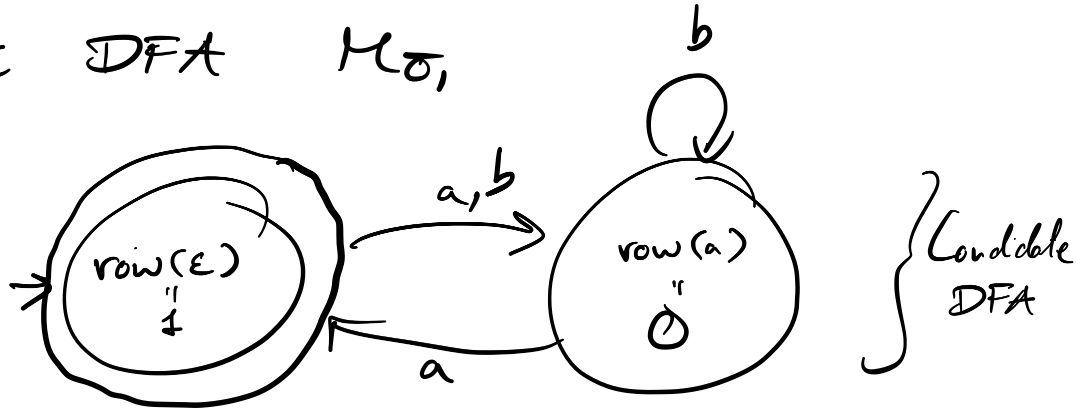
$S \cdot \Sigma$

		ε
S	ε	1
	a	0
$S\cdot\Sigma$	b	0
	a a	1
	a b	0

Is $\mathcal{O}_1$ closed? Yes

Is $\mathcal{O}_1$ consistent? Yes (vacuously)

Create DFA $M_{\mathcal{O}_1}$



$\hookrightarrow$ $T \to$ NO, $bb \in L$

How do I integrate counterexamples from T?

$w = \sigma_1 \sigma_2 \cdots \sigma_n$

counter-example string

Then $S \leftarrow S \cup \underbrace{\text{prefixes}}_{= \{\varepsilon, b, bb\}}(bb)$

2. $\mathcal{O}_2$ where $S = \{\varepsilon, a, b, bb\}$ $E = \{\varepsilon\}$

	ε
ε	1
a	0
b	0
$\boxed{bb}$	1
aa	1
ab	0
$\boxed{ba}$	0
$\boxed{bba}$	0
$\boxed{bbb}$	0

NEW!

Is $\mathcal{O}_2$ closed?
Yes

Is $\mathcal{O}_2$ consistent?
No! $row(a \cdot a \cdot \varepsilon) = 1$
but $row(b \cdot a \cdot \varepsilon) = 0$

Add $a \cdot \varepsilon$ to E, $E \leftarrow E \cup \{a \cdot \varepsilon\}$

3. Create $\mathcal{O}_3$ S from $\mathcal{O}_2$
$E = \{\varepsilon, a\}$

$$\underbrace{\qquad\qquad\qquad\qquad}_{}^{\mathcal{O}_3}$$

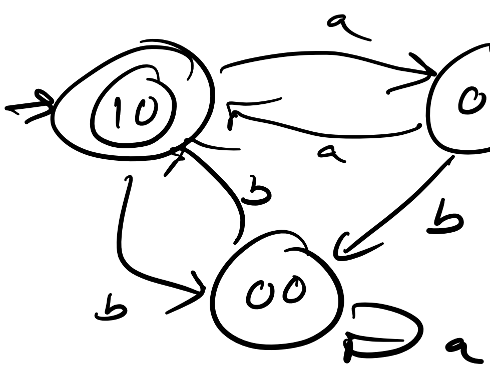
	ε	a
ε	1	0
a	0	1
b	0	0
bb	1	0
aa	1	0
ab	0	0
ba	0	0
bba	0	1
bbb	0	0

$T(aa \cdot a)$

Is σ_3 closed? Yes

Consistent? Yes

M_{σ_3} → Candidate DFA



→ T → No

↓

abb $\notin L$

$S \leftarrow S \cup$ prefixes (abb)

$\{\varepsilon, a, ab, abb\}$

1. $S \leftarrow S \cup$ prefixes(abb)
$E \leftarrow \{\varepsilon, a\}$

$\mathcal{O}_1$

	ε	a	
ε	1	0	
a	0	1	
b	0	0	
bb	1	0	
ab	0	0	
abb	0	1	
aa	1	0	
ba	0	0	
bba	0	1	
bbb	0	0	
abba	1	0	
abbb	0	0	
aba	0	0	

Is $\mathcal{O}_1$ closed? Yes

Is $\mathcal{O}_1$ consistent? Inconsistent

b/c row(b) = 00 row(ab) = 00
but row(bb) = 10 $\neq$ row(abb) = 01

↳ The first place that rows differ is @ $T(\boxed{b}\,b\,\varepsilon) \neq T(\boxed{ab}\,b\,\varepsilon)$

initial string from S

$\sigma \in \Sigma$

final distinguishing string

Add the string $b \cdot \varepsilon$ to E

$$E \leftarrow E \cup \{b \cdot \varepsilon\}$$

5. Construct Θ_5

	ε	a	b
ε	1	0	0
a	0	1	0
b	0	0	1
bb	1	0	0
ab	0	0	0
abb	0	1	0
aa	1	0	0
ba	0	0	0
bba	0	1	0
bbb	0	0	1
abba	1	0	0
abbb	0	0	0
aba	0	0	1

Closed!

Consistent!