Finite Automata

Alphabets and Languages

Def

$\Sigma^0 = \{\varepsilon\}$ contains only the empty word

Σ - alphabet (finite set)

$\Sigma^n = \{a_1 a_2 .. a_n \mid a_i \in \Sigma\}$ is the set of words of length n

$\Sigma^* = \{w \mid \exists n \in \mathbb{N}.\ \exists\, a_1, a_2, .., a_n \in \Sigma.\ w = a_1 a_2 .. a_n\}$ is the set of all words over Σ

A language L over Σ is a subset $L \subseteq \Sigma^*$

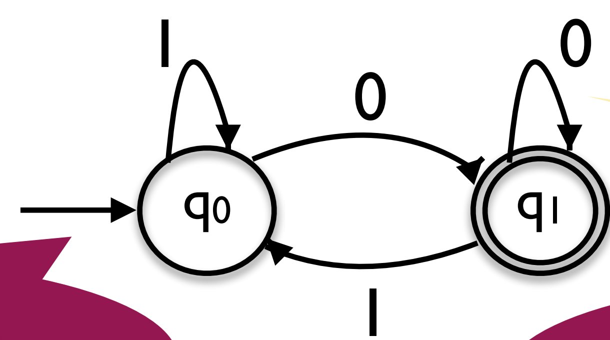
Deterministic Automata (DFA)

alphabet

Informal example

$\Sigma = \{0,1\}$

M_1:



q_0, q_1 are states

transitions, labelled by alphabet symbols

q_0 is initial

q_1 is final

Accepts the language $L(M_1) = \{w \in \Sigma^* \mid w \text{ ends with a } 0\} = \Sigma^*0$

regular language

regular expression

DFA

Definition

A deterministic automaton M is a tuple $M = (Q, \Sigma, \delta, q_0, F)$ where

Q is a finite set of states
Σ is a finite alphabet
$\delta: Q \times \Sigma \longrightarrow Q$ is the transition function
q_0 is the initial state, $q_0 \in Q$
F is a set of final states, $F \subseteq Q$

In the example M_1

$M_1 = (Q, \Sigma, \delta, q_0, F)$ for

$Q = \{q_0, q_1\}$ $F = \{q_1\}$

$\Sigma = \{0, 1\}$

$\delta(q_0, 0) = q_1, \delta(q_0, 1) = q_0$
$\delta(q_1, 0) = q_1, \delta(q_1, 1) = q_0$

DFA

The extended transition function

Given $M = (Q, \Sigma, \delta, q_0, F)$ we can extend $\delta: Q \times \Sigma \longrightarrow Q$ to

$\delta^*: Q \times \Sigma^* \longrightarrow Q$

In M_1, $\delta^*(q_0, 110010) = q_1$

inductively, by:

$\delta^*(q, \varepsilon) = q$ and $\delta^*(q,wa) = \delta(\delta^*(q,w), a)$

Definition

$L(M_1) = \{w0 | w \in \{0,1\}^*\}$

The language recognised / accepted by a deterministic finite automaton $M = (Q, \Sigma, \delta, q_0, F)$ is

$$L(M) = \{w \in \Sigma^* | \; \delta^*(q_0,w) \in F\}$$

Regular languages and operations

$L(M_1) = \{w0 | w \in \{0,1\}^*\}$ is regular

Definition

Let Σ be an alphabet. A language L over Σ ($L \subseteq \Sigma^*$) is regular iff it is recognised by a DFA.

Regular operations

Let L, L_1, L_2 be languages over Σ. Then $L_1 \cup L_2$, $L_1 \cdot L_2$, and L^* are languages, where

$$L_1 \cdot L_2 = \{w_1 \cdot w_2 \mid w_1 \in L_1, w_2 \in L_2\}$$

$$L^* = \{w \mid \exists n \in \mathbb{N}.\ \exists\ w_1, w_2, .., w_n \in L.\ w = w_1 w_2 .. w_n\}$$

$\varepsilon \in L^*$ always

Closure under regular operations

also under intersection

Theorem C1

The class of regular languages is closed under union

We can already prove these!

Theorem C2

The class of regular languages is closed under complement

Theorem C3

The class of regular languages is closed under concatenation

But not yet these two…

Theorem C4

The class of regular languages is closed under Kleene star finite representation of infinite languages

Regular expressions

inductive

example: $(ab \cup a)^*$

Definition

Let Σ be an alphabet. The following are regular expressions

1. a for $a \in \Sigma$
2. ε
3. $\varnothing$
4. $(R_1 \cup R_2)$ for R_1, R_2 regular expressions
5. $(R_1 \cdot R_2)$ for R_1, R_2 regular expressions
6. $(R_1)^*$ for R_1 regular expression

corresponding languages

$$L(a) = \{a\}$$
$$L(\varepsilon) = \{\varepsilon\}$$
$$L(\varnothing) = \varnothing$$
$$L(R_1 \cup R_2) = L(R_1) \cup L(R_2)$$
$$L(R_1 \cdot R_2) = L(R_1) \cdot L(R_2)$$
$$L(R_1^*) = L(R_1)^*$$

Equivalence of regular expressions and regular languages

Theorem (Kleene)

A language is regular (i.e., recognised by a finite automaton) iff it is the language of a regular expression.

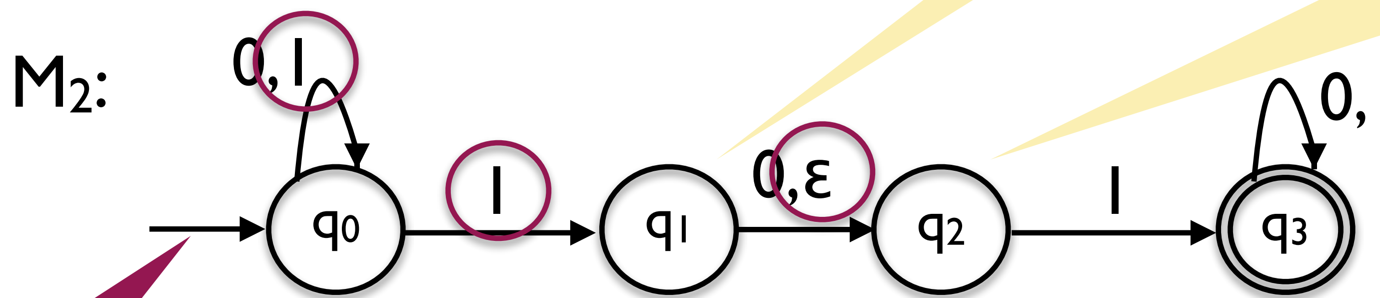
needs nondeterminism

Proof $\Leftarrow$ easy, as the constructions for the closure properties,
$\Rightarrow$ not so easy, we'll skip it for now…

Nondeterministic Automata (NFA)

Informal example

$\Sigma = \{0,1\}$



no 1 transition

no 0 transition

sources of nondeterminism

Accepts a word iff there exists an accepting run

NFA

Definition

A nondeterministic automaton M is a tuple $M = (Q, \Sigma, \delta, q_0, F)$ where

Q is a finite set of states
Σ is a finite alphabet
$\delta: Q \times \Sigma_\varepsilon \longrightarrow \mathcal{P}(Q)$ is the transition function
q_0 is the initial state, $q_0 \in Q$
F is a set of final states, $F \subseteq Q$

$\Sigma_\varepsilon = \Sigma \cup \{\varepsilon\}$

In the example M_2

$M_2 = (Q, \Sigma, \delta, q_0, F)$ for

$Q = \{q_0, q_1, q_2, q_3\}$

$\Sigma = \{0, 1\}$ $F = \{q_3\}$

$\delta(q_0, 0) = \{q_0\}$
$\delta(q_0, 1) = \{q_0, q_1\}$
$\delta(q_0, \varepsilon) = \varnothing$
…..

ε-closure of q, all states reachable by ε-transitions from q

NFA

$E(q) = \{q' \mid q' = q \vee \exists n \in \mathbb{N}^+ . \exists q_0, .., q_n \in Q . q_0 = q, q_n = q', q_{i+1} \in \delta(q_i, \varepsilon), \text{ for } i = 0, .., n-1\}$

The extended transition function

Given an NFA $M = (Q, \Sigma, \delta, q_0, F)$ we can extend $\delta: Q \times \Sigma_\varepsilon \longrightarrow \mathcal{P}(Q)$ to

$$\delta^*: Q \times \Sigma^* \longrightarrow \mathcal{P}(Q)$$

$E(X) = \bigcup_{x \in X} E(x)$

In M_2, $\delta^*(q_0, 0110) = \{q_0, q_2, q_3\}$

inductively, by:

$$\delta^*(q, \varepsilon) = E(q) \quad \text{and} \quad \delta^*(q, wa) = E\left(\bigcup_{q' \in \delta^*(q,w)} \delta(q', a)\right)$$

Definition

The language recognised / accepted by a ... automaton $M = (Q, \Sigma, \delta, q_0, F)$ is

$L(M_2) = \{u101w \mid u,w \in \{0,1\}^*\}$
$\cup$
$\{u11w \mid u,w \in \{0,1\}^*\}$

$$L(M) = \{w \in \Sigma^* \mid \delta^*(q_0, w) \cap F \neq \varnothing\}$$

Equivalence of automata

Definition

Two automata M_1 and M_2 are equivalent if $L(M_1) = L(M_2)$

Theorem NFA ~ DFA

Every NFA has an equivalent DFA

Proof via the "powerset construction" / determinization

Corollary

A language is regular iff it is recognised by a NFA

Closure under regular operations

Theorem C1

The class of regular languages is closed under union

Theorem C2

The class of regular languages is closed under complement

Theorem C3

The class of regular languages is closed under concatenation

Now we can prove these too

Theorem C4

The class of regular languages is closed under Kleene star

Nonregular languages

every long enough word of a regular language can be pumped

Theorem (Pumping Lemma)

If L is a regular language, then there is a number $p \in \mathbb{N}$ (the pumping length) such that for any $w \in L$ with $|w| \geq p$, there exist $x, y, z \in \Sigma^*$ such that $w = xyz$ and

1. $xy^iz \in L$, for all $i \in \mathbb{N}$
2. $|y| > 0$
3. $|xy| \leq p$

Proof easy, using the pigeonhole principle

Example "corollary"

$L = \{ 0^n 1^n \mid n \in \mathbb{N}\}$ is nonregular.

Note the logical structure!