Automata and Logic

Aart Middeldorp and Johannes Niederhauser

Outline

Definitions

- nondeterministic finite automaton (NFA) is quintuple $N = (Q, \Sigma, \Delta, S, F)$ with

  ① Q: finite set of states

  ② Σ: input alphabet

  ③ $\Delta \colon Q \times \Sigma \to 2^Q$: transition function

  ④ $S \subseteq Q$: set of start states

  ⑤ $F \subseteq Q$: final (accept) states

- $\widehat{\Delta} \colon 2^Q \times \Sigma^* \to 2^Q$ is inductively defined by

$$\widehat{\Delta}(A, \epsilon) = A \qquad\qquad \widehat{\Delta}(A, xa) = \bigcup_{q \in \widehat{\Delta}(A,x)} \Delta(q, a)$$

- $x \in \Sigma^*$ is accepted by N if $\widehat{\Delta}(S, x) \cap F \neq \varnothing$

Theorem

every set accepted by NFA is regular

Definitions

- NFA with ϵ-transitions (NFA_ϵ) is sextuple $N = (Q, \Sigma, \epsilon, \Delta, S, F)$ such that
  1. $\epsilon \notin \Sigma$
  2. $M_\epsilon = (Q, \Sigma \cup \{\epsilon\}, \Delta, S, F)$ is NFA over alphabet $\Sigma \cup \{\epsilon\}$
- ϵ-closure of set $A \subseteq Q$ is defined as $C_\epsilon(A) = \bigcup \{ \widehat{\Delta}_{N_\epsilon}(A, x) \mid x \in \{\epsilon\}^* \}$
- $\widehat{\Delta}_N \colon 2^Q \times \Sigma^* \to 2^Q$ is inductively defined by

$$\widehat{\Delta}_N(A, \epsilon) = C_\epsilon(A) \qquad \widehat{\Delta}_N(A, xa) = \bigcup \{ C_\epsilon(\Delta(q, a)) \mid q \in \widehat{\Delta}_N(A, x) \}$$

Theorem

- every set accepted by NFA_ϵ is regular
- regular sets are effectively closed under union, concatenation, and asterate

Automata

- (deterministic, non-deterministic, alternating) finite automata
- regular expressions
- (alternating) Büchi automata

Logic

- (weak) monadic second-order logic
- Presburger arithmetic
- linear-time temporal logic

Outline

Definitions

- regular expression α over alphabet Σ:

$$a \in \Sigma \qquad \epsilon \qquad \varnothing \qquad \beta + \gamma \qquad \beta\gamma \qquad \beta^*$$

- set of strings $L(\alpha) \subseteq \Sigma^*$ matched by regular expression α:

$$L(a) = \{a\} \qquad L(\beta + \gamma) = L(\beta) \cup L(\gamma)$$
$$L(\epsilon) = \{\epsilon\} \qquad L(\beta\gamma) = L(\beta)L(\gamma)$$
$$L(\varnothing) = \varnothing \qquad L(\beta^*) = L(\beta)^*$$

Example

regular expression $(a + b)^*b$ matches all strings over $\Sigma = \{a, b\}$ that end with b

Definition

regular expressions α and β are equivalent $(\alpha \equiv \beta)$ if $L(\alpha) = L(\beta)$

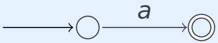
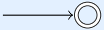
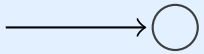
Theorem

finite automata and regular expressions are equivalent:

$$\text{for all } A \subseteq \Sigma^* \quad A \text{ is regular} \quad \Longleftrightarrow \quad A = L(\alpha) \text{ for some regular expression } \alpha$$

Proof ($\Longleftarrow$)

induction on regular expression α

α	$L(\alpha)$	finite automaton
$a \in \Sigma$	$\{a\}$	
ϵ	$\{\epsilon\}$	
$\varnothing$	$\varnothing$	

α	$L(\alpha)$
$\beta + \gamma$	$L(\beta) \cup L(\gamma)$
$\beta\gamma$	$L(\beta)L(\gamma)$
β^*	$L(\beta)^*$

$L(\beta)$ and $L(\gamma)$ are regular according to induction hypothesis

$\Longrightarrow$ $L(\alpha)$ is regular according to closure properties of regular sets

Lemma (Arden's Lemma)

if $A, B, X \subseteq \Sigma^*$ such that $X = AX \cup B$ and $\epsilon \notin A$ then $X = A^*B$

Proof

$X \subseteq A^*B$

- let $x \in X = AX \cup B$
- induction on $|x|$
  - $x \in AX \implies x = ay$ for some $a \in A$ and $y \in X \implies y \in A^*B \implies x \in A^*B$
  - $x \in B \implies x \in A^*B$ because $\epsilon \in A^*$

$X \supseteq A^*B$

- $x \in A^*B \implies x = x_1 \cdots x_k y$ for some $x_1, \dots, x_k \in A$ and $y \in B$
- induction on k
  - $k = 0 \implies x = y \in B \subseteq X$
  - $k > 0 \implies x_2 \cdots x_k y \in X \implies x \in AX \subseteq X$

Theorem

finite automata and regular expressions are equivalent:

$$\text{for all } A \subseteq \Sigma^* \quad A \text{ is regular} \quad \Longleftrightarrow \quad A = L(\alpha) \text{ for some regular expression } \alpha$$

Proof ($\Longrightarrow$)

given NFA $N = (Q, \Sigma, \Delta, S, F)$ with $Q = \{1, \dots, n\}$ and $S = \{1\}$

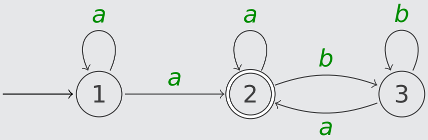
- define system of equations

$$X_i = \left(\bigcup_{a \in \Sigma} \; \bigcup_{j \in \Delta(i,a)} \{a\} X_j \right) \cup \begin{cases} \{\epsilon\} & \text{if } i \in F \\ \varnothing & \text{otherwise} \end{cases}$$

  with unknowns $X_1, \dots, X_n$

- transform X_1 into regular expression by successive substitution and Arden's lemma

Example



$X_1 = a\,X_1 + a\,X_2$	$X_2 = a\,X_2 + b\,X_3 + \epsilon$	$X_3 = a\,X_2 + b\,X_3$	
$X_1 = a^*a\,X_2$	$X_2 = a^*(b\,X_3 + \epsilon)$	$X_3 = b^*a\,X_2$	(Arden's lemma)
$X_1 = a^*a\,X_2$	$X_2 = a^*(bb^*a\,X_2 + \epsilon)$		(substitute)
$X_1 = a^*a\,X_2$	$X_2 = a^*bb^*a\,X_2 + a^*$		(distribute)
$X_1 = a^*a\,X_2$	$X_2 = (a^*bb^*a)^*a^*$		(Arden's lemma)
$X_1 = a^*a(a^*bb^*a)^*a^*$			(substitute)

Outline

Question

Which of the following strings belong to $L((aba + ab + b)^*)$?

A ϵ

B $ababa$

C all strings over $\{a, b\}$ that start with b

D all strings over $\{a, b\}$ that do not contain two consecutive b's

Outline

Theorem

regular sets are effectively closed under homomorphic image and preimage

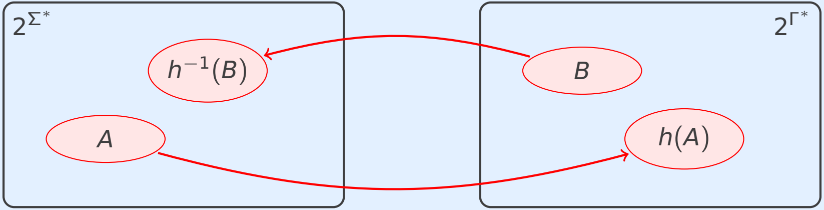
Definitions

- homomorphism is mapping $h\colon \Sigma^* \to \Gamma^*$ such that

$$h(\epsilon) = \epsilon \qquad\qquad h(xy) = h(x)h(y)$$

  so homomorphism is completely determined by its effect on Σ

- if $A \subseteq \Sigma^*$ then $h(A) = \{h(x) \mid x \in A\} \subseteq \Gamma^*$ "image of A under h"
- if $B \subseteq \Gamma^*$ then $h^{-1}(B) = \{x \mid h(x) \in B\} \subseteq \Sigma^*$ "preimage of B under h"



- homomorphism $h\colon \Sigma^* \to \Gamma^*$
- $h^{-1}(h(A)) \supseteq A$
- $h(h^{-1}(B)) \subseteq B$

Example

$\Sigma = \Gamma = \{0,1\} \quad h(0) = 11 \quad h(1) = 1 \quad A = B = \{0\}$

- $h^{-1}(h(A)) = h^{-1}(\{11\}) = \{0,11\} \supsetneq A$
- $h(h^{-1}(B)) = h(\varnothing) = \varnothing \subsetneq B$

Example

$A \subseteq \{0,1\}^*$ is regular $\implies$ $\{xy \mid x1y \in A\}$ is regular

- $\Sigma = \{0,1\}$ and $\Gamma = \{0,1,2\}$
- define homomorphisms $h, i \colon \Gamma^* \to \Sigma^*$ by

$$h(0) = 0 \qquad h(1) = h(2) = 1 \qquad i(0) = 0 \qquad i(1) = 1 \qquad i(2) = \epsilon$$

- $h^{-1}(A) = \{x \mid h(x) \in A\}$
- $h^{-1}(A) \cap L((0+1)^*2(0+1)^*) = \{x2y \mid x1y \in A\}$
- $\{xy \mid x1y \in A\} = i(h^{-1}(A) \cap L((0+1)^*2(0+1)^*))$ is regular

Theorem

regular sets are effectively closed under homomorphic image and preimage

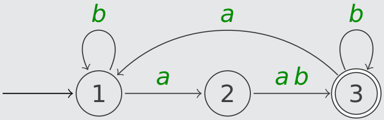
Proof

- NFA $M = (Q, \Gamma, \Delta, S, F)$
- homomorphism $h\colon \Sigma^* \to \Gamma^*$
- $h^{-1}(L(M)) = L(M')$ for NFA $M' = (Q, \Sigma, \Delta', S, F)$ with $\Delta'(q, a) = \widehat{\Delta}(\{q\}, h(a))$
- claim: $\widehat{\Delta'}(A, x) = \widehat{\Delta}(A, h(x))$ for all $A \subseteq Q$ and $x \in \Sigma^*$

  proof of claim: easy induction on $|x|$

Example

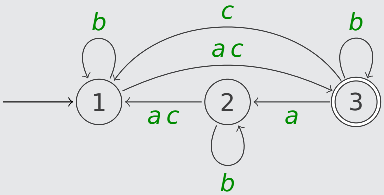
- DFA M



- homomorphism $h\colon \{a, b, c\}^* \to \{a, b\}^*$

$$h(a) = aa \qquad h(b) = \epsilon \qquad h(c) = bab$$

- DFA M'



$$\delta'(3, c) = \widehat{\delta}(3, bab) = 1$$

Theorem

regular sets are effectively closed under homomorphic image and preimage

Proof

- regular expression α over Σ
- homomorphism $h\colon \Sigma^* \to \Gamma^*$
- $h(L(\alpha)) = L(\alpha')$ for regular expression α' defined inductively:

$$a' = h(a) \quad \text{for } a \in \Sigma \qquad\qquad (\beta + \gamma)' = \beta' + \gamma'$$

$$\epsilon' = \epsilon \qquad\qquad (\beta\gamma)' = \beta'\gamma'$$

$$\varnothing' = \varnothing \qquad\qquad \beta^{*\prime} = \beta'^{*}$$

Definitions

- Hamming distance $H(x,y)$ is number of places where bit strings x and y differ
- if $|x| \neq |y|$ then $H(x,y) = \infty$
- $N_k(A) = \{x \in \{0,1\}^* \mid H(x,y) \leqslant k \text{ for some } y \in A\}$

Lemma

$A \subseteq \{0,1\}^*$ is regular $\implies$ $\forall\, k \in \mathbb{N}$ $N_k(A)$ is regular

Proof

$$\begin{aligned} D_k &= \{x \in (\{0,1\} \times \{0,1\})^* \mid x \text{ contains at most } k \text{ pairs } (0,1) \text{ or } (1,0)\} \qquad \text{is regular} \\ &= \{x \in (\{0,1\} \times \{0,1\})^* \mid H(\mathsf{fst}(x), \mathsf{snd}(x)) \leqslant k\} \end{aligned}$$

$$N_k(A) = \mathsf{fst}(\mathsf{snd}^{-1}(A) \cap D_k)$$

Example

- $A = \{0011\} \quad k = 2$
- $N_k(A)$ consists of

0011	1011	0111	0001	0010	1111
1001	1010	0101	0110	0000	

- $\mathrm{fst}(\mathrm{snd}^{-1}(A) \cap D_k)$ consists of

0000	0001	0010	0011	0100	0101
0011	0011	0011	0011	0011	0011
0110	0111	1000	1001	1010	1011
0011	0011	0011	0011	0011	0011
1100	1101	1110	1111		
0011	0011	0011	0011		

Outline

Remark

most decision problems concerning regular sets are decidable

Theorem

problems

instance: DFA M and string x	instance: DFA M	instance: DFAs M and N
question: $x \in L(M)$?	question: $L(M) = \varnothing$?	question: $L(M) = L(N)$?

are decidable

Remark

representation of regular sets (DFA, NFA, regular expression) may affect complexity of decision problems

Outline

Kozen

- Lecture 7 – 10

Important Concepts

- Arden's lemma
- homomorphism
- homomorphic image
- homomorphic preimage
- regular expression

homework for October 25