Automata on Finite Words

Definition

A *non-deterministic finite automaton* (NFA) over Σ is a tuple $A = \langle S, I, T, F \rangle$ where:

- S is a finite set of *states*,
- $I \subseteq S$ is a set of *initial states*,
- $T \subseteq S \times \Sigma \times S$ is a *transition relation*,
- $F \subseteq S$ is a set of *final states*.

We denote $T(s, \alpha) = \{s' \in S \mid (s, \alpha, s') \in T\}$. When T is clear from the context we denote $(s, \alpha, s') \in T$ by $s \xrightarrow{\alpha} s'$.

Determinism and Completeness

Definition 1 *An automaton* $A = \langle S, I, T, F \rangle$ *is* deterministic *(DFA) iff* $\|I\| = 1$ *and, for each* $s \in S$ *and for each* $\alpha \in \Sigma$, $\|T(s, \alpha)\| \leq 1$.

If A is deterministic we write $T(s, \alpha) = s'$ instead of $T(s, \alpha) = \{s'\}$.

Definition 2 *An automaton* $A = \langle S, I, T, F \rangle$ *is* complete *iff for each* $s \in S$ *and for each* $\alpha \in \Sigma$, $\|T(s, \alpha)\| \geq 1$.

Runs and Acceptance Conditions

Given a finite word $w \in \Sigma^*$, $w = \alpha_1\alpha_2\ldots\alpha_n$, a *run* of A over w is a finite sequence of states $s_1, s_2, \ldots, s_n, s_{n+1}$ such that $s_1 \in I$ and $s_i \xrightarrow{\alpha_i} s_{i+1}$ for all $1 \leq i \leq n$.

A run over w between s_i and s_j is denoted as $s_i \xrightarrow{w} s_j$.

The run is said to be *accepting* iff $s_{n+1} \in F$. If A has an accepting run over w, then we say that A *accepts* w.

The language of A, denoted $\mathcal{L}(A)$ is the set of all words accepted by A.

A set of words $S \subseteq \Sigma^*$ is *recognizable* if there exists an automaton A such that $S = \mathcal{L}(A)$.

Determinism, Completeness, again

Proposition 1 *If A is deterministic, then it has at most one run for each input word.*

Proposition 2 *If A is complete, then it has at least one run for each input word.*

Determinization

Theorem 1 *For every NFA A there exists a DFA A_d such that* $\mathcal{L}(A) = \mathcal{L}(A_d)$.

Let $A_d = \langle 2^S, \{I\}, T_d, \{G \subseteq S \mid G \cap F \neq \emptyset\}\rangle$, where

$$(S_1, \alpha, S_2) \in T_d \iff S_2 = \{s' \mid \exists s \in S_1 \,.\, (s, \alpha, s') \in T\}$$

This definition is known as subset construction

On the Exponential Blowup of Complementation

Theorem 2 *For every $n \in \mathbb{N}$, $n \geq 1$, there exists an automaton A, with $size(A) = n + 1$ such that no deterministic automaton with less than 2^n states recognizes the complement of $\mathcal{L}(A)$.*

Let $\Sigma = \{a, b\}$ and $L = \{uav \mid u, v \in \Sigma^*, |v| = n - 1\}$.

There exists a NFA with exactly $n + 1$ states which recognizes L.

Suppose that $B = \langle S, \{s_0\}, T, F\rangle$, is a (complete) DFA with $\|S\| < 2^n$ that accepts $\Sigma^* \setminus L$.

On the Exponential Blowup of Complementation

$\|\{w \in \Sigma^* \mid |w| = n\}\| = 2^n$ and $\|S\| < 2^n$ (by the pigeonhole principle)

$\Rightarrow \exists uav_1, ubv_2 \ . \ |uav_1| = |ubv_2| = n$ and $s \in S \ . \ s_0 \xrightarrow{uav_1} s$ and $s_0 \xrightarrow{ubv_2} s$

Let s_1 be the (unique) state of B such that $s \xrightarrow{u} s_1$.

Since $|uav_1| = n$, then $uav_1u \in L \Rightarrow uav_1u \notin \mathcal{L}(B)$, i.e. s is not accepting.

On the other hand, $ubv_2u \notin L \Rightarrow ubv_2u \in \mathcal{L}(B)$, i.e. s is accepting, contradiction.

Completion

Lemma 1 *For every NFA A there exists a complete NFA A_c such that $\mathcal{L}(A) = \mathcal{L}(A_c)$.*

Let $A_c = \langle S \cup \{\sigma\}, I, T_c, F\rangle$, where $\sigma \notin S$ is a new sink state. The transition relation T_c is defined as:

$$\forall s \in S \forall \alpha \in \Sigma \ . \ (s, \alpha, \sigma) \in T_c \iff \forall s' \in S \ . \ (s, \alpha, s') \notin T$$

and $\forall \alpha \in \Sigma \ . \ (\sigma, \alpha, \sigma) \in T_c$.

Closure Properties

Theorem 3 *Let* $A_1 = \langle S_1, I_1, T_1, F_1\rangle$ *and* $A_2 = \langle S_2, I_2, T_2, F_2\rangle$ *be two NFA. There exists automata* $\bar{A}_1$, $A_\cup$ *and* $A_\cap$ *that recognize the languages* $\Sigma^* \setminus \mathcal{L}(A_1)$, $\mathcal{L}(A_1) \cup \mathcal{L}(A_2)$, *and* $\mathcal{L}(A_1) \cap \mathcal{L}(A_2)$ *respectivelly.*

Let $A' = \langle S', I', T', F'\rangle$ be the complete deterministic automaton such that $\mathcal{L}(A_1) = \mathcal{L}(A')$, and $\bar{A}_1 = \langle S', I', T', S' \setminus F'\rangle$.

Let $A_\cup = \langle S_1 \cup S_2, I_1 \cup I_2, T_1 \cup T_2, F_1 \cup F_2\rangle$.

Let $A_\cap = \langle S_1 \times S_2, I_1 \times I_2, T_\cap, F_1 \times F_2\rangle$ where:

$$(\langle s_1, t_1\rangle, \alpha, \langle s_2, t_2\rangle) \in T_\cap \iff (s_1, \alpha, s_2) \in T_1 \text{ and } (t_1, \alpha, t_2) \in T_2$$

Projections

Let the input alphabet $\Sigma = \Sigma_1 \times \Sigma_2$. Any word $w \in \Sigma^*$ can be uniquely identified to a pair $\langle w_1, w_2 \rangle \in \Sigma_1^* \times \Sigma_2^*$ such that $|w_1| = |w_2| = |w|$.

The *projection* operations are
$pr_1(L) = \{u \in \Sigma_1^* \mid \langle u, v \rangle \in L, \text{ for some } v \in \Sigma_2^*\}$ and
$pr_2(L) = \{v \in \Sigma_2^* \mid \langle u, v \rangle \in L, \text{ for some } u \in \Sigma_1^*\}$.

Theorem 4 *If the language $L \subseteq (\Sigma_1 \times \Sigma_2)^*$ is recognizable, then so are the projections $pr_i(L)$, for $i = 1, 2$.*

Remark

The operations of union, intersection and complement correspond to the boolean $\vee$, $\wedge$ and $\neg$.

The projection corresponds to the first-order existential quantifier $\exists x$.

The Myhill-Nerode Theorem

Let $A = \langle S, I, T, F \rangle$ be an automaton over the alphabet Σ^*.

Define the relation $\sim_A \subseteq \Sigma^* \times \Sigma^*$ as:

$$u \sim_A v \iff [\forall s, s' \in S \;.\; s \xrightarrow{u} s' \iff s \xrightarrow{v} s']$$

$\sim_A$ is an equivalence relation of finite index

Let $L \subseteq \Sigma^*$ be a language. Define the relation $\sim_L \subseteq \Sigma^* \times \Sigma^*$ as:

$$u \sim_L v \iff [\forall w \in \Sigma^* \;.\; uw \in L \iff vw \in L]$$

$\sim_L$ is an equivalence relation

The Myhill-Nerode Theorem

Theorem 5 *A language* $L \subseteq \Sigma^*$ *is recognizable iff* $\sim_L$ *is of finite index.*

"$\Rightarrow$" Suppose $L = \mathcal{L}(A)$ for some automaton A.

$\sim_A$ is of finite index.

for all $u, v \in \Sigma^*$ we have $u \sim_A v \Rightarrow u \sim_L v$

index of $\sim_L \leq$ index of $\sim_A < \infty$

The Myhill-Nerode Theorem

"$\Leftarrow$" $\sim_L$ is an equivalence relation of finite index, and let $[u]$ denote the equivalence class of $u \in \Sigma^*$.

$A = \langle S, I, T, F \rangle$, where:

- $S = \{[u] \mid u \in \Sigma^*\}$,
- $I = [\epsilon]$,
- $[u] \xrightarrow{\alpha} [v] \iff u\alpha \sim_L v$,
- $F = \{[u] \mid u \in L\}$.

Isomorphism and Canonical Automata

Two automata $A_i = \langle S_i, I_i, T_i, F_i \rangle$, $i = 1, 2$ are said to be *isomorphic* iff there exists a bijection $h : S_1 \to S_2$ such that, for all $s, s' \in S_1$ and for all $\alpha \in \Sigma$ we have :

- $s \in I_1 \iff h(s) \in I_2$,
- $(s, \alpha, s') \in T_1 \iff (h(s), \alpha, h(s')) \in T_2$,
- $s \in F_1 \iff h(s) \in F_2$.

For DFA all minimal automata are isomorphic.

For NFA there may be more non-isomorphic minimal automata.

Pumping Lemma

Lemma 2 (Pumping) *Let* $A = \langle S, I, T, F\rangle$ *be a finite automaton with* $size(A) = n$*, and* $w \in \mathcal{L}(A)$ *be a word of length* $|w| \geq n$*. Then there exists three words* $u, v, t \in \Sigma^*$ *such that:*

1. $|v| \geq 1$,
2. $w = uvt$ *and,*
3. *for all* $k \geq 0$, $uv^k t \in \mathcal{L}(A)$.

Example

$L = \{a^n b^n \mid n \in \mathbb{N}\}$ is not recognizable:

Suppose that there exists an automaton A with $size(A) = N$, such that $L = \mathcal{L}(A)$.

Consider the word $a^N b^N \in L = \mathcal{L}(A)$.

There exists words u, v, w such that $|v| \geq 1$, $uvw = a^N b^N$ and $uv^k w \in L$ for all $k \geq 1$.

- $v = a^m$, for some $m \in \mathbb{N}$.
- $v = a^m b^p$ for some $m, p \in \mathbb{N}$.
- $v = b^m$, for some $m \in \mathbb{N}$.

Decidability

Given automata A and B:

- **Emptiness** $\mathcal{L}(A) = \emptyset$?
- **Equality** $\mathcal{L}(A) = \mathcal{L}(B)$?
- **Infinity** $\|\mathcal{L}(A)\| < \infty$?
- **Universality** $\mathcal{L}(A) = \Sigma^*$?

Emptiness

Theorem 6 *Let A be an automaton with size$(A) = n$. If $\mathcal{L}(A) \neq \emptyset$, then there exists a word of length less than n that is accepted by A.*

Let u be the shortest word in $\mathcal{L}(A)$.

If $|u| < n$ we are done.

If $|u| \geq n$, there exists $u_1, v, u_2 \in \Sigma^*$ such that $|v| > 1$ and $u_1 v u_2 = u$.

Then $u_1 u_2 \in \mathcal{L}(A)$ and $|u_1 u_2| < |u_1 v u_2|$, contradiction.

Everything is decidable

Theorem 7 *The emptiness, equality, infinity and universality problems are decidable for automata on finite words.*

Automata on Finite Words and WS1S

WS1S

Let $\Sigma = \{a, b, \ldots\}$ be a finite alphabet.

Any finite word $w \in \Sigma^*$ induces the *finite* sets $p_a = \{p \mid w(p) = a\}$.

- $x \leq y$: x is less than y,
- $s(x) = y$: y is the successor of x,
- $p_a(x)$: a occurs at position x in w

Remember that $\leq$ and $s(.)$ can be defined one from another.

Problem Statement

Let $\mathcal{L}(\varphi) = \{w \mid \mathfrak{m}_w \models \varphi\}$

A language $L \subseteq \Sigma^*$ is said to be WS1S-*definable* iff there exists a WS1S formula φ such that $L = \mathcal{L}(\varphi)$.

1. Given A build φ_A such that $\mathcal{L}(A) = \mathcal{L}(\varphi)$
2. Given φ build A_φ such that $\mathcal{L}(A) = \mathcal{L}(\varphi)$

The recognizable and WS1S-definable languages coincide

Coding of Σ

Let $m \in \mathbb{N}$ be the smallest number such that $\|\Sigma\| \leq 2^m$.

W.l.o.g. assume that $\Sigma = \{0, 1\}^m$, and let $X_1 \dots X_p, x_{p+1}, \dots x_m$

A word $w \in \Sigma^*$ induces an *interpretation* of $X_1 \dots X_p, x_{p+1}, \dots x_m$:

- $i \in I_w(X_j)$ iff the j-th element of w_i is 1, and
- $I_w(x_j) = i$ iff w_i has 1 on the j-th position and, for all $k \neq i$ w_k has 0 on the j-th position.

Example

Example 1 *Let* $\Sigma = \{a, b, c, d\}$*, encoded as* $a = (00)$*,* $b = (01)$*,* $c = (10)$ *and* $d = (11)$*. Then the word abbaacdd induces the valuation* $X_1 = \{5, 6, 7\}$*,* $X_2 = \{1, 2, 6, 7\}$*.* $\Box$

From Automata to Formulae

Let $A = \langle S, I, T, F \rangle$ with $S = \{s_1, \ldots, s_p\}$, and $\Sigma = \{0, 1\}^m$.

Build $\Phi_A(X_1, \ldots, X_m)$ such that $\forall w \in \Sigma^* \, . \, w \in \mathcal{L}(A) \iff w \models \Phi_A$

Let $a \in \{0, 1\}^m$. Let $\Phi_a(x, X_1, \ldots, X_m)$ be the conjunction of:

- $X_i(x)$ if the $a_i = 1$, and
- $\neg X_i(x)$ otherwise.

For all $w \in \Sigma^*$ we have $w \models \forall x \, . \, \bigvee_{a \in \Sigma} \Phi_a(x, \vec{X})$

Notice that $\Phi_a \wedge \Phi_b$ is unsatisfiable, for $a \neq b$.

Coding of S

Let $\{Y_0, \ldots, Y_p\}$ be set variables.

Y_i is the set of all positions labeled by A with state s_i during some run

$$\Phi_S(Y_1, \ldots, Y_p) \ : \ \forall z \ . \bigvee_{1 \leq i \leq p} Y_i(z) \ \wedge \bigwedge_{1 \leq i < j \leq p} \neg \exists z \ . \ Y_i(z) \wedge Y_j(z)$$

Coding of I

Every run starts from an initial state:

$$\Phi_I(Y_1, \ldots, Y_p) \;:\; \exists x \forall y \,.\, x \leq y \wedge \bigvee_{s_i \in I} Y_i(x)$$

Coding of T

Consider the transition $s_i \xrightarrow{a} s_j$:

$$\Phi_T(X_1, \ldots, X_m, Y_1, \ldots, Y_p) \; : \; \forall x \, . \, x \neq s(x) \wedge Y_i(x) \wedge \Phi_a(x, \vec{X}) \rightarrow \bigvee_{(s_i, a, s_j) \in T} Y_j(s(x))$$

Coding of F

The last state on the run is a final state:

$$\Phi_F(Y_1, \ldots, Y_p) \; : \; \exists x \forall y \, . \, y \leq x \wedge \bigvee_{s_i \in F} Y_i(x)$$

$$\Phi_A = \exists Y_1 \ldots \exists Y_p \, . \, \Phi_S \wedge \Phi_I \wedge \Phi_T \wedge \Phi_F$$

From Formulae to Automata

Let $\Phi(X_1, \ldots, X_p, x_{p+1}, \ldots, x_m)$ be a WS1S formula.

We build an automaton A_Φ such that $\mathcal{L}(A) = \mathcal{L}(\Phi)$.

Let $\Phi(X_1, X_2, x_3, x_4)$ be:

1. $X_1(x_3)$
2. $x_3 \leq x_4$
3. $X_1 = X_2$

From Formulae to Automata

A_Φ is built by induction on the structure of Φ:

- for $\Phi = \phi_1 \wedge \phi_2$ we have $\mathcal{L}(A_\Phi) = \mathcal{L}(A_{\phi_1}) \cap \mathcal{L}(A_{\phi_2})$
- for $\Phi = \phi_1 \vee \phi_2$ we have $\mathcal{L}(A_\Phi) = \mathcal{L}(A_{\phi_1}) \cup \mathcal{L}(A_{\phi_2})$
- for $\Phi = \neg\phi$ we have $\mathcal{L}(A_\Phi) = \overline{\mathcal{L}(A_\phi)}$
- for $\Phi = \exists X_i \, . \, \phi$, we have $\mathcal{L}(A_\Phi) = pr_i(\mathcal{L}(A_\phi))$.

Consequences

Theorem 8 *A language $L \subseteq \Sigma^*$ is definable in WS1S iff it is recognizable.*

Corollary 1 *The SAT problem for WS1S is decidable.*

Lemma 3 *Any WS1S formula $\phi(X_1, \ldots, X_m)$ is equivalent to an WS1S formula of the form $\exists Y_1 \ldots \exists Y_p \,.\, \varphi$, where φ does not contain other set variables than $X_1, \ldots, X_m, Y_1, \ldots, Y_p$.*

Finite Semigroups

Monoids and Semigroups

A *semigroup* is a set $(\mathcal{S}, \cdot)$, where:

$$\forall x, y, z \in \mathcal{S} \ . \ (x \cdot y) \cdot z = x \cdot (y \cdot z)$$

A *monoid* is a semigroup $(\mathcal{M}, \cdot)$ such that

$$\exists 1 \in \mathcal{M} \ . \ x \cdot 1 = 1 \cdot x = x$$

If S is a semigroup, we denote by S^1 the monoid obtained by adding an identity element to S, i.e. $S^1 = S \cup \{1\}$.

Example 2

- Σ^+ *with concatenation forms a semigroup*
- Σ^* *with concatenation and* ϵ *forms a monoid*
- *the set* $\mathcal{R}(S)$ *of relations over a set* S *with composition of relations and the identity relation forms a monoid*

Morphisms

Let $(S, \cdot)$ and $(T, \times)$ be semigroup(s). A *semigroup morphism* is a function $\varphi : S \to T$ such that :

$$\forall x, y \in S \ . \ \varphi(x \cdot y) = \varphi(x) \times \varphi(y)$$

Let $(S, \cdot, 1_S)$ and $(T, \times, 1_T)$ be monoid(s). A *monoid morphism* is a semigroup morphism $\varphi : S \to T$ such that :

$$\varphi(1_S) = 1_T$$

Recognition by Morphism

Let S and T be semigroup(s). A surjective semigroup morphism $\varphi : S \to T$ recognizes $I \subseteq S$ iff there exists $J \subseteq T$ such that:

$$I = \varphi^{-1}(J)$$

Example 3 *Let S be the semigroup of multiplicative matrices generated by:*

$$A = \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix} \quad B = \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix}$$

$\Sigma = \{a, b\}$ *and* $\varphi : \Sigma \to S$ *defined by* $\varphi(a) = A$ *and* $\varphi(b) = B$*. Then*

$\varphi^{-1}(A) = (ab)^*a,\ \varphi^{-1}(B) = (ba)^*b,\ \varphi^{-1}(AB) = (ab)^+,\ \varphi^{-1}(BA) = (ba)^+$

$$\varphi^{-1}(AA) = \varphi^{-1}(BB) = \Sigma^*aa\Sigma^* \cup \Sigma^*bb\Sigma^*$$

From Automata to Semigroups

Let $A = \langle S, I, T, F \rangle$ be an automaton over an alphabet Σ. Define the semigroup morphism $\varphi : \Sigma^+ \to \mathcal{R}(S)$ by

$$\varphi(\alpha) = \{(s, s') \in S \times S \mid (s, \alpha, s') \in T\}$$

$$\begin{aligned} \varphi(\alpha_0\alpha_1 \ldots \alpha_n) &= \varphi(\alpha_0) \circ \varphi(\alpha_1) \circ \ldots \circ \varphi(\alpha_n) \\ &= \{(s, s') \in S \times S \mid s \xrightarrow{\alpha_0\alpha_1\ldots\alpha_n} s'\} \end{aligned}$$

$\mathcal{L}(A) = \varphi^{-1}(\{(s, s') \in I \times F \mid \text{there exists a path from } s \text{ to } s'\})$

$\varphi(\Sigma^+)$ is called the *transition semigroup* of A

The *transition monoid* of A, is $\varphi(\Sigma^*)$, where φ is a monoid morphism

From Semigroups to Automata

Let $(S, \cdot)$ be a semigroup and $\varphi : \Sigma^+ \to S$ be a surjective semigroup morphism recognizing $L \subseteq \Sigma^+$, i.e. $L = \varphi^{-1}(F)$ for some $F \subseteq S$.

Then $A = \langle S^1, \{1\}, T, F \rangle$, where

$$T = \{(s, \alpha, s \cdot \varphi(\alpha)) \mid s \in S, \alpha \in \Sigma\}$$

recognizes L.

$$\begin{aligned} \mathcal{L}(A) &= \{u \in \Sigma^+ \mid 1 \cdot \varphi(u) \in F\} \\ &= \{u \in \Sigma^+ \mid \varphi(u) \in F\} \\ &= \varphi^{-1}(F) \\ &= L \end{aligned}$$

Congruences

Definition 3 *An equivalence relation* $R \subseteq \Sigma^* \times \Sigma^*$ *is said to be a* left-congruence *iff for all* $u, v, w \in \Sigma^*$ *we have* $u \cong v \Rightarrow wu \cong wv$*.*

Definition 4 *An equivalence relation* $R \subseteq \Sigma^* \times \Sigma^*$ *is said to be a* right-congruence *iff for all* $u, v, w \in \Sigma^*$ *we have* $u \; R \; v \Rightarrow uw \; R \; vw$*.*

Definition 5 *An equivalence relation* $R \subseteq \Sigma^* \times \Sigma^*$ *is said to be a* congruence *iff it is both a left- and a right-congruence.*

Ex: the Myhill-Nerode equivalence $\sim_L$ is a right-congruence.

The Syntactic Semigroup (Monoid)

Let $L \subseteq \Sigma^+$ (Σ^*) be a language. The relation $\simeq_L \subseteq \Sigma^+ \times \Sigma^+$ $(\Sigma^* \times \Sigma^*)$ is defined as

$$x \simeq_L y \ : \ \forall u, v \in \Sigma^* \ . \ uxv \in L \iff uyv \in L$$

The quotient semigroup (monoid) $\Sigma^+_{/\simeq_L}$ $(\Sigma^*_{/\simeq_L})$ is called the *syntactic monoid* of L.

It is the coarsest semigroup (monoid) that recognizes L. It is also the transition semigroup (monoid) of the minimal deterministic automaton recognizing L.

A language is recognizable iff it is recognized by a finite semigroup.

Syntactic Monoid Example

Let $L = (ab)^*$. Then we have:

$$\begin{array}{ccccccc} \epsilon & \simeq_L & ab & \simeq_L & abab & \simeq_L & ababab \ldots \\ a & \simeq_L & aba & \simeq_L & ababa & \simeq_L & abababa \ldots \\ b & \simeq_L & bab & \simeq_L & babab & \simeq_L & bababab \ldots \end{array}$$

$$\begin{array}{ccccccccc} \epsilon & \not\simeq_L & a & a & \not\simeq_L & ab & a & \not\simeq_L & b \\ \epsilon & \not\simeq_L & b & b & \not\simeq_L & ba & ab & \not\simeq_L & ba \end{array}$$

$$\ldots$$

Syntactic Monoid Example

Let S be the semigroup of multiplicative matrices generated by:

$$A = \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix} \quad B = \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix}$$

$\Sigma = \{a, b\}$ and $\varphi : \Sigma \to S$ defined by $\varphi(a) = A$ and $\varphi(b) = B$.