Automata theory and its applications

Lecture 5-6: Chomsky hierarchy-Regular languages

Zhilin Wu

State Key Laboratory of Computer Science,
Institute of Software, Chinese Academy of Sciences

November 4, 2012

Outline

Nondeterministic finite state automata (NFA)

A nondeterministic finite state automaton $\mathcal{A}$ is a tuple $(Q, \Sigma, \delta, q_0, F)$ such that

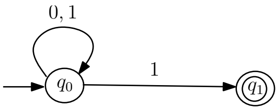
- Q is a finite set of states,
- Σ is the finite alphabet,
- q_0 is the initial state,
- $F \subseteq Q$ is the set of final states,
- $\delta \subseteq Q \times \Sigma \times Q$ is the transition relation.

A *run* of $\mathcal{A}$ over a word $w = a_1 \ldots a_n \in \Sigma^*$ is a state sequence $q_0 q_1 \ldots q_n$ such that for every $i : 1 \leqslant i \leqslant n$, $(q_{i-1}, a_i, q_i) \in \delta$.

A run $q_0 \ldots q_n$ is *accepting* if $q_n \in F$.

If there is an accepting run of $\mathcal{A}$ over w, then w is said to be accepted by $\mathcal{A}$.

Let $\mathcal{L}(\mathcal{A})$ denote the set of words accepted by $\mathcal{A}$.



NFA with ε transitions (ε-NFA)

A ε-NFA is a NFA $(Q, \Sigma, \delta, q_0, F)$ with

$$\delta \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times Q.$$

A *run* of $\mathcal{A}$ over a word $w \in \Sigma^*$ is a sequence $q_0 a_1 q_1 a_2 \ldots a_n q_n$ s.t.

for every $i : 1 \leqslant i \leqslant n$, $(q_{i-1}, a_i, q_i) \in \delta$ *and* $a_1 \ldots a_n = w$.

Note that a_i $(1 \leqslant i \leqslant n)$ may be ε.

Proposition. ε-NFA $\equiv$ NFA.

Because a NFA is a ε-NFA, it suffices to prove that

from any ε-NFA $\mathcal{A}$, an equivalent NFA $\mathcal{B}$ can be constructed.

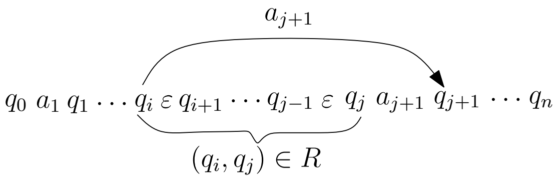
NFA with ε transitions (ε-NFA): continued

Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$.
Compute inductively the ε-reachability relation R as follows.

- $R_0 = \{(q,q) \mid q \in Q\} \cup \{(q,q') \mid (q,\varepsilon,q') \in \delta\}$,
- $R_{i+1} = R_i \cup \{(q,q') \mid \exists q'' \in Q, (q,q''), (q'',q') \in R_i\}$.

Then let $\mathcal{B} = (Q, \Sigma, \delta', q_0, F)$ such that

$$\delta' = \{(q,a,q') \mid a \in \Sigma, \exists q''.(q,q'') \in R, (q'',a,q') \in \delta\}.$$



Right linear grammar

A right linear grammar is a grammar $G = (\mathcal{N}, \Sigma, \mathcal{P}, S)$ such that $\mathcal{P}$ contains rules of the form

$$A \to aB, \text{ or } A \to a, \text{ or } A \to \varepsilon, \text{ where } a \in \Sigma, B \in \mathcal{N}.$$

Proposition. Right linear grammar $\equiv$ NFA

From right linear grammar to NFA

Let $G = (\mathcal{N}, \Sigma, \mathcal{P}, S)$ be a right linear grammar.
Construct $\mathcal{A} = (\mathcal{N} \cup \{q_f\}, \Sigma, \delta, S, F)$ as follows.

- $F = \{q_f\} \cup \{A \mid A \to \varepsilon\}$.
- δ is defined by the following rules,

$$(A, a, B) \in \delta \textit{ iff } A \to aB, \quad (A, a, q_f) \in \delta \textit{ iff } A \to a.$$

Right linear grammar

A right linear grammar is a grammar $G = (\mathcal{N}, \Sigma, \mathcal{P}, S)$ such that $\mathcal{P}$ contains rules of the form

$$A \to aB, \text{ or } A \to a, \text{ or } A \to \varepsilon, \text{ where } a \in \Sigma, B \in \mathcal{N}.$$

Proposition. Right linear grammar $\equiv$ NFA

From right linear grammar to NFA

Let $G = (\mathcal{N}, \Sigma, \mathcal{P}, S)$ be a right linear grammar.
Construct $\mathcal{A} = (\mathcal{N} \cup \{q_f\}, \Sigma, \delta, S, F)$ as follows.

- $F = \{q_f\} \cup \{A \mid A \to \varepsilon\}$.
- δ is defined by the following rules,

$$(A, a, B) \in \delta \text{ iff } A \to aB, \quad (A, a, q_f) \in \delta \text{ iff } A \to a.$$

From NFA to right linear grammar

Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a NFA.
Construct a right linear grammar $G = (Q, \Sigma, \mathcal{P}, q_0)$ as follows.

$$q \to aq' \text{ iff } (q, a, q') \in \delta, \quad q \to \varepsilon \text{ iff } q \in F.$$

Regular expression

Syntax of regular expressions is defined by the following rules,

$$r := a \mid \varepsilon \mid r_1 \cup r_2 \mid r_1 \cdot r_2 \mid r_1^*.$$

Semantics,

- $L(a) = \{a\}$, $L(\varepsilon) = \{\varepsilon\}$,
- $L(r_1 \cup r_2) = L(r_1) \cup L(r_2)$,
- $L(r_1 \cdot r_2) = L(r_1) \cdot L(r_2)$,
- $L(r_1^*) = (L(r_1))^*$,

where

- $L_1 \cdot L_2 = \{uv \mid u \in L_1, v \in L_2\}$,
- $L^* = \bigcup_{i=1}^{\infty} L^i$,
- $L^0 = \{\varepsilon\}$, $L^i = L^{i-1} \cdot L$ for any $i > 0$.

Equivalence between regular expression and NFA

Theorem. Regular expression $\equiv$ NFA.

From regular expression to NFA

Let r be a regular expression. Construct $\mathcal{A}$ from r inductively as follows.

- If $r = a$, then $\mathcal{A} = (\{q_0, q_1\}, \Sigma, \delta, q_0, \{q_1\})$ such that $(q_0, a, q_1) \in \delta$.
- If $r = \varepsilon$, then $\mathcal{A} = (\{q_0\}, \Sigma, \varnothing, q_0, \{q_0\})$,
- if $r = r_1 \cup r_2$, let $\mathcal{A}_i = (Q_i, \Sigma, \delta_i, q_0^i, F_i)$ for r_i (where $i = 1, 2$), then $\mathcal{A} = (Q_1 \cup Q_2 \cup \{q_0\}, \Sigma, \delta, q_0, F)$ such that
  - $\delta = \delta_1 \cup \delta_2 \cup \{(q_0, a, q) \mid (q_0^1, a, q) \in \delta_1 \text{ or } (q_0^2, a, q) \in \delta_2\}$,
  - $F = F_1 \cup F_2$.
- If $r = r_1 \cdot r_2$, let $\mathcal{A}_i = (Q_i, \Sigma, \delta_i, q_0^i, F_i)$ for r_i (where $i = 1, 2$), then $\mathcal{A} = (Q_1 \cup Q_2, \Sigma, \delta, q_0^1, F)$ such that
  - $\delta = \delta_1 \cup \delta_2 \cup \{(q, a, q') \mid q \in F_1, (q_0^2, a, q') \in \delta_2\}$,
  - if $q_0^1 \in F_1$ and $q_0^2 \in F$, then $F = F_2 \cup \{q_0^1\}$, otherwise $F = F_2$.
- If $r := (r_1)^*$, let $\mathcal{A}_1 = (Q_1, \Sigma, \delta_1, q_0^1, F_1)$ for r_1, then $\mathcal{A} = (Q_1 \cup \{q_0\}, \Sigma, \delta, q_0, \{q_0\})$ such that $\delta = \delta_1 \cup \{(q_0, a, q) \mid (q_0^1, a, q) \in \delta_1\} \cup \{(q, a, q_0) \mid \exists q' \in F_1.(q, a, q') \in \delta_1\}$.

Equivalence between regular expression and NFA

Theorem. Regular expression $\equiv$ NFA.

From NFA to regular expression

Let $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$ be a NFA. Suppose $Q = \{q_1, \ldots, q_n\}$.
Consider the languages R_{ijk}'s:

> *The set of words over which there is a run from q_i to q_j such that the indices of all the intermediate states (not including q_i, q_j) are $\leqslant k$.*

Define r_{ijk}'s inductively for R_{ijk}'s as follows.

- $r_{ij0} = \bigcup\limits_{a:(q_i,a,q_j)\in\delta} a.$
- $r_{ijk} = r_{ij(k-1)} \cup r_{ik(k-1)}(r_{kk(k-1)})^* r_{kj(k-1)}.$

Let $r = \bigcup\limits_{q_i \in F} r_{1in}.$

Claim. $L(r) = L(\mathcal{A})$.

Regular languages:

> *Languages that can be defined by NFA (or right linear grammar or regular expression).*

Outline

Pumping lemma

Pumping lemma. Let L be a regular language. Then there is $n \geqslant 1$ such that for any $u \in L$ with $|u| \geqslant n$, u can be decomposed into xyz satisfying that

- $|y| \geqslant 1$, $|y| \leqslant n$,
- and for any $i \in \mathbb{N}$, $xy^iz \in L$.

Suppose $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ is an NFA defining L.
Let $n = |Q|$.
Then for any $u = a_1 \ldots a_m \in L$ such that $m \geqslant n$,

consider an accepting run of $\mathcal{A}$ on u, say $q_0 a_1 q_1 \ldots a_m q_m$.

Because $m \geqslant n$, it follows that

there must be $r, s : 0 \leqslant r < s \leqslant m$ such that $q_r = q_s$.

Let $x = a_1 \ldots a_r$, $y = a_{r+1} \ldots a_s$, $z = a_{s+1} \ldots a_m$.
For any $i \in \mathbb{N}$,

$$q_0 a_1 q_1 \ldots a_r (q_r a_{r+1} \ldots q_{s-1} a_s)^i q_s a_{s+1} q_{s+1} \ldots a_m q_m$$

is still an accepting run of $\mathcal{A}$. Therefore, $xy^iz \in L$.

Applications of pumping lemma

$L = \{a^n b^n \mid n \geqslant 1\}$ is not regular.

To the contrary, suppose that L is regular. Then there is a $n \geqslant 1$ satisfying the condition in pumping lemma.

Consider $u = a^n b^n$. Then $u = xyz$ such that $|y| \geqslant 1$ and $|y| \leqslant n$ and $xy^i z \in L$ for any $i \in \mathbb{N}$.

- $y = a^i$ or $y = b^i$: $xy^2 z \notin L$.
- $y = a^i b^j$: $xy^2 z \notin L$.

In both situations, we have a contradiction.

Corollary. Regular language $\subset$ CFL.

Outline

Myhill-Nerode equivalence relation

Let $L \subseteq \Sigma^*$. Define $\sim_L$ over Σ^* as follows: For any $u, v \in \Sigma^*$,

$$u \sim_L v \text{ iff } \forall z \in \Sigma^*. uz \in L \Leftrightarrow vz \in L.$$

Proposition. $\sim_L$ is a right congruence.

Proof.

$\sim_L$ is an equivalence relation: reflexive, transitive and symmetric.
For any $u, v : u \sim_L v$ and $z \in \Sigma^*$, $uz \sim_L vz$. □

The index of $\sim_L$ is the number of equivalence classes of $\sim_L$.

Example: Let $L = a^*b$, then $\sim_L$ contains

three equivalence classes a^*, a^*b, $a^*b(a \cup b)^+$.

Myhill-Nerode theorem

Theorem. Let $L \subseteq \Sigma^*$. Then L is regular iff $\sim_L$ is of finite index.

"Only If" direction

Suppose L is defined by the NFA $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$.
Then for any $x \in \Sigma^*$, define $R(x)$ as

the set of states that can be reached from q_0 after reading x.

It follows that

for any $u, v \in \Sigma^*$, $R(u) = R(v) \Rightarrow u \sim_L v$.

Therefore, $\sim_L$ is of finite index.

"If" direction

Suppose $\sim_L$ is of finite index.
Let $E_1, \ldots, E_n$ be the equivalence classes of $\sim_L$ and $E_1 = [\varepsilon]$.
Then $\mathcal{A} = (\{E_1, \ldots, E_n\}, \Sigma, \delta, E_1, \{E_i \mid E_i \cap L \neq \emptyset\})$, where

$$(E_i, a, E_j) \in \delta \text{ iff } \exists u \in E_i . ua \in E_j.$$

Outline

DFA and subset construction

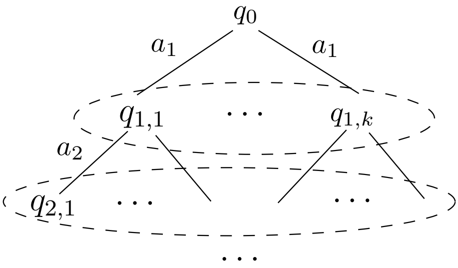
A DFA is an NFA $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ such that for every $q \in Q$, $a \in \Sigma$, $|\delta(q, a)| \leqslant 1$.

A *complete* DFA is a DFA such that for every $q \in Q$, $a \in \Sigma$, $|\delta(q, a)| = 1$.

Theorem. NFA $\equiv$ DFA.

It suffices to show that for every NFA $\mathcal{A}$, an equivalent DFA $\mathcal{B}$ can be constructed.

The intuition:



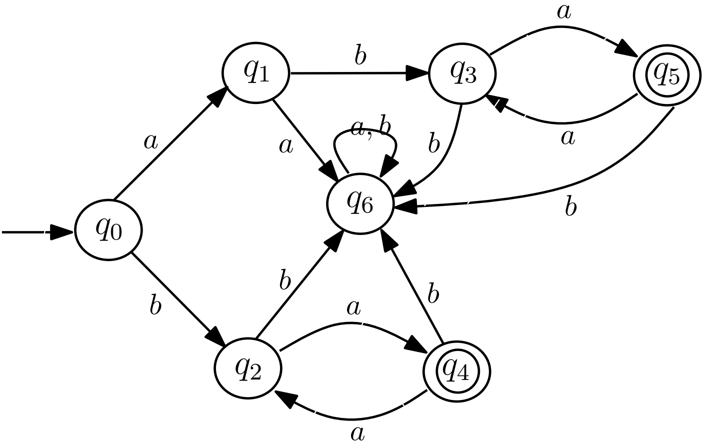
DFA and subset construction: continued

Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$, construct $\mathcal{B} = (Q', \Sigma, \delta', q_0', F')$ as follows.

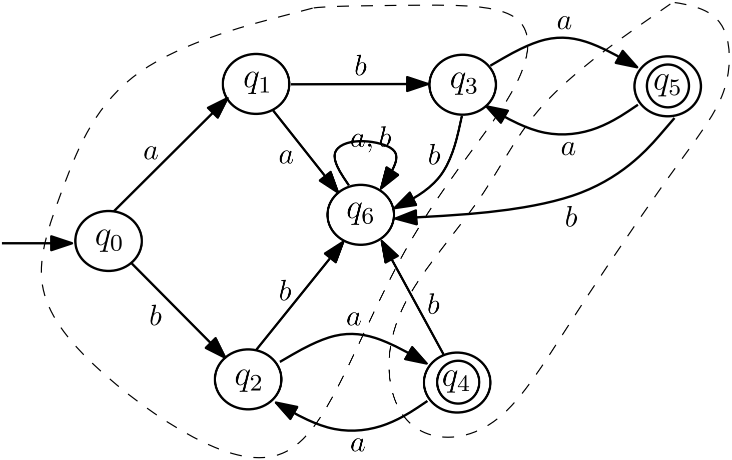
- $Q' = Pow(Q)$,
- $q_0' = \{q_0\}$,
- $F' = \{X \in Pow(Q) \mid X \cap F \neq \emptyset\}$,
- for every $X \in Pow(Q)$, $\delta(X, a) = \bigcup\limits_{q \in X} \delta(q, a)$.

The above construction is called the subset construction.

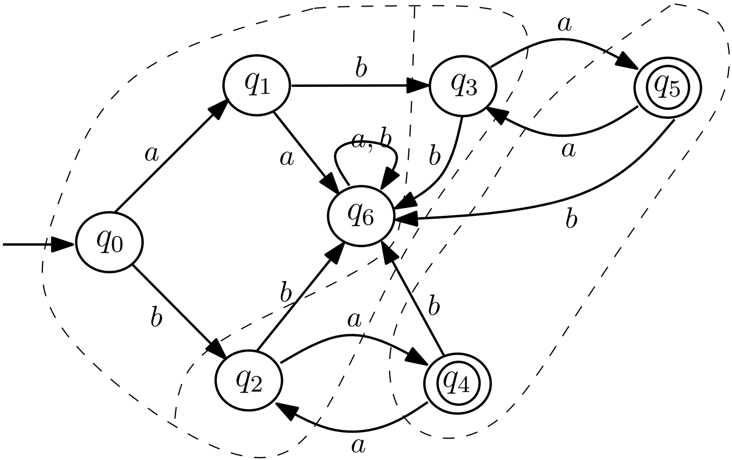
Minimization: An example



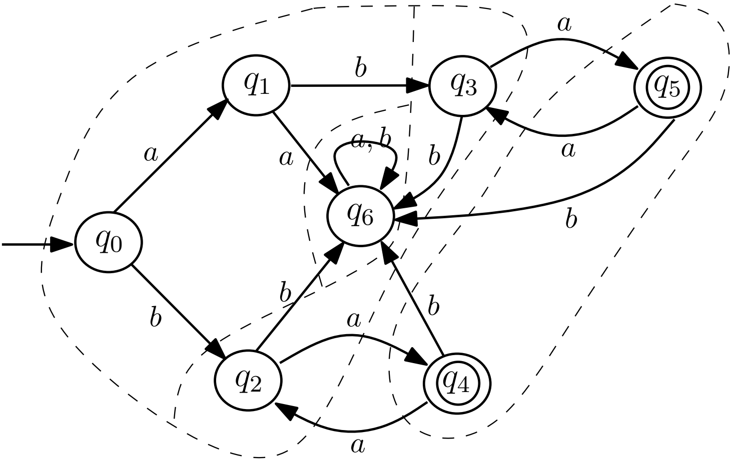
Minimization: An example



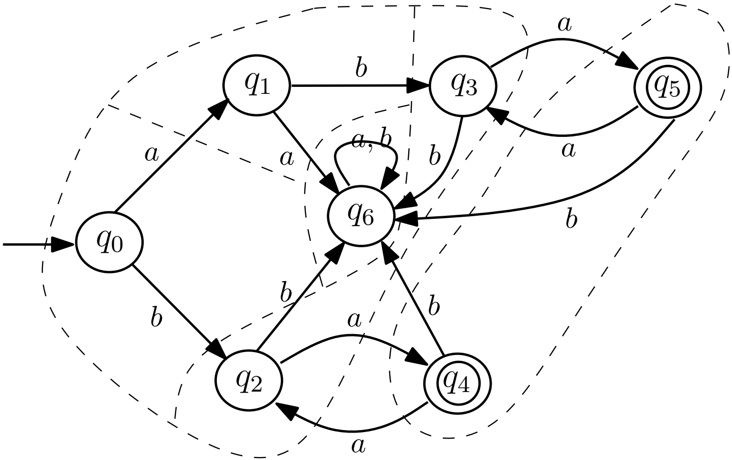
Minimization: An example



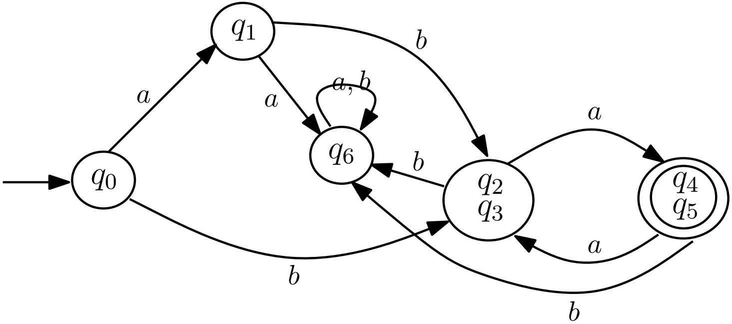
Minimization: An example



Minimization: An example



Minimization: An example



Uniqueness of minimum-size DFA

Theorem. For every regular language $L \subseteq \Sigma^*$,

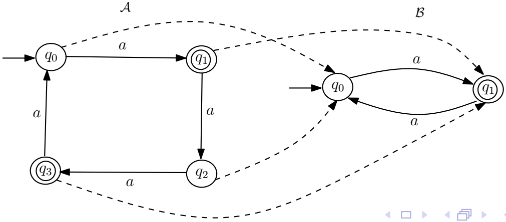
there is a unique complete DFA of the minimum size.

Morphism

Let $\mathcal{A} = (Q_1, \Sigma, \delta_1, q_0^1, F_1)$ and $\mathcal{B} = (Q_2, \Sigma, \delta_2, q_0^2, F_2)$ be two DFAs.
A morphism from $\mathcal{A}$ to $\mathcal{B}$ is a surjective mapping h from Q_1 to Q_2 such that

- $h(q_0^1) = q_0^2$,
- for every $q \in Q$, $q \in F_1$ iff $h(q) \in F_2$,
- for every $q, q' \in Q, a \in \Sigma$ s.t. $\delta_1(q, a) = q'$, it holds $\delta_2(h(q), a) = h(q')$.

A morphism is called an isomorphism iff h is bijective.



Uniqueness of minimum-size DFA: continued

Let $L \subseteq \Sigma^*$ be a regular language.

Let $\mathcal{A}_L = (Q_L, \Sigma, \delta_L, q_0^L, F_L)$ be the DFA corresponding to $\sim_L$,

- Q_L is the set of equivalence classes of $\sim_L$,
- $\delta_L([x], a) = [xa]$ for any $x \in \Sigma^*$,
- $q_0^L = [\varepsilon]$, $F_L = \{[x] \mid x \in L\}$.

Claim. For every DFA $\mathcal{A}$ such that $L(\mathcal{A}) = L$, $\exists$ a morphism from $\mathcal{A}$ to $\mathcal{A}_L$.

Proof.

Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a complete DFA such that $L(\mathcal{A}) = L$.
Then for every $x, y \in \Sigma^*$ such that $\delta(q_0, x) = \delta(q_0, y)$, we have

for every $z \in \Sigma^*$, $xz \in L$ iff $yz \in L$.

Define a mapping $h : Q \to Q_L$ as follows: $h(q) = [x]$ with $\delta(q_0, x) = q$.

- h is surjective since $\mathcal{A}$ is complete,
- $h(q_0) = [\varepsilon] = q_0^L$,
- if $q \in F$, then $h(q) = [x]$ for $x \in \Sigma^*$ s.t. $\delta(q_0, x) = q$. So $x \in L$, $[x] \in F_L$,
- if $\delta(q, a) = q'$, then $\delta_L(h(q), a) = \delta_L([x], a) = [xa] = h(q')$ with $\delta(q_0, x) = q$.

□

Uniqueness of minimum-size DFA: continued

Let $L \subseteq \Sigma^*$ be a regular language.
Let $\mathcal{A}_L = (Q_L, \Sigma, \delta_L, q_0^L, F_L)$ be the DFA corresponding to $\sim_L$,

- Q_L is the set of equivalence classes of $\sim_L$,
- $\delta_L([x], a) = [xa]$ for any $x \in \Sigma^*$,
- $q_0^L = [\varepsilon]$, $F_L = \{[x] \mid x \in L\}$.

Claim. For every DFA $\mathcal{A}$ such that $L(\mathcal{A}) = L$, $\exists$ a morphism from $\mathcal{A}$ to $\mathcal{A}_L$.

Because $|\mathcal{A}| \geqslant |\mathcal{A}_L|$,

it follows that $\mathcal{A}_L$ is the DFA defining L of the minimum size.

Uniqueness.

Suppose $\mathcal{B}$ is a DFA of the minimum size defining L.

Then $\mathcal{B}$ has the same size as $\mathcal{A}_L$.

According to the claim, there is a morphism h from $\mathcal{B}$ to $\mathcal{A}_L$.

Because h is surjective, it follows that h is bijective.

Therefore, h is an isomorphism from $\mathcal{B}$ to $\mathcal{A}_L$.

Minimization

The problem

Given a DFA $\mathcal{A}$, construct an equivalent DFA $\mathcal{B}$ of the minimum size.

Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a DFA.

Compute inductively an equivalence relation $\approx_{\mathcal{A}}$ over Q as follows, until $\approx^i_{\mathcal{A}} = \approx^{i+1}_{\mathcal{A}}$.

- $q \approx^0_{\mathcal{A}} q'$ iff $q \in F \Leftrightarrow q' \in F$,
- $q \approx^{i+1}_{\mathcal{A}} q'$ iff $q \approx^i_{\mathcal{A}} q'$ and $\forall a \in \Sigma$, $\delta(q, a) \approx^i_{\mathcal{A}} \delta(q', a)$.

Because $\forall i.\ \approx^{i+1}_{\mathcal{A}} \subseteq \approx^i_{\mathcal{A}}$, it follows that the above procedure terminates.

Minimization

The problem

Given a DFA $\mathcal{A}$, construct an equivalent DFA $\mathcal{B}$ of the minimum size.

Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a DFA.

Compute inductively an equivalence relation $\approx_{\mathcal{A}}$ over Q as follows, until $\approx^i_{\mathcal{A}} = \approx^{i+1}_{\mathcal{A}}$.

- $q \approx^0_{\mathcal{A}} q'$ iff $q \in F \Leftrightarrow q' \in F$,
- $q \approx^{i+1}_{\mathcal{A}} q'$ iff $q \approx^i_{\mathcal{A}} q'$ and $\forall a \in \Sigma$, $\delta(q, a) \approx^i_{\mathcal{A}} \delta(q', a)$.

Because $\forall i.\ \approx^{i+1}_{\mathcal{A}} \subseteq \approx^i_{\mathcal{A}}$, it follows that the above procedure terminates.

Observation. $\approx_{\mathcal{A}}$ enjoys the following properties.

- $q \approx_{\mathcal{A}} q' \Rightarrow q \in F$ iff $q' \in F$ and $\forall x \in \Sigma^*$, $\delta(q, x) \approx_{\mathcal{A}} \delta(q', x)$,
- $q \not\approx_{\mathcal{A}} q' \Rightarrow \exists x \in \Sigma^*$ s.t. $\delta(q, x) \in F$ and $\delta(q', x) \notin F$ or vice versa.

Minimization

The problem

Given a DFA $\mathcal{A}$, construct an equivalent DFA $\mathcal{B}$ of the minimum size.

Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a DFA.

Compute inductively an equivalence relation $\approx_{\mathcal{A}}$ over Q as follows, until $\approx^i_{\mathcal{A}} = \approx^{i+1}_{\mathcal{A}}$.

- $q \approx^0_{\mathcal{A}} q'$ iff $q \in F \Leftrightarrow q' \in F$,
- $q \approx^{i+1}_{\mathcal{A}} q'$ iff $q \approx^i_{\mathcal{A}} q'$ and $\forall a \in \Sigma$, $\delta(q, a) \approx^i_{\mathcal{A}} \delta(q', a)$.

Because $\forall i. \approx^{i+1}_{\mathcal{A}} \subseteq \approx^i_{\mathcal{A}}$, it follows that the above procedure terminates.

Observation. $\approx_{\mathcal{A}}$ enjoys the following properties.

- $q \approx_{\mathcal{A}} q' \Rightarrow q \in F$ iff $q' \in F$ and $\forall x \in \Sigma^*$, $\delta(q, x) \approx_{\mathcal{A}} \delta(q', x)$,
- $q \not\approx_{\mathcal{A}} q' \Rightarrow \exists x \in \Sigma^*$ s.t. $\delta(q, x) \in F$ and $\delta(q', x) \notin F$ or vice versa.

Corollary. $\forall u, v \in \Sigma^*$, $u \sim_L v$ iff $\delta(q_0, u) \approx_{\mathcal{A}} \delta(q_0, v)$.

Therefore,

$\mathcal{A}/\approx_{\mathcal{A}}$ *is the DFA of the minimum size.*

Outline

Boolean operations

Theorem. Regular languages are closed under union, intersection and complementation.

Union:

Let $\mathcal{A}_1 = (Q_1, \Sigma, \delta_1, q_0^1, F_1)$ and $\mathcal{A}_2 = (Q_2, \Sigma, \delta_2, q_0^2, F_2)$ be two NFAs.
Then $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ defines $L(\mathcal{A}_1) \cup L(\mathcal{A}_2)$, where

- $Q = Q_1 \cup Q_2 \cup \{q_0\}$, $F = F_1 \cup F_2$,
- $\delta = \delta_1 \cup \delta_2 \cup \left\{(q_0, a, q) \mid (q_0^1, a, q) \in \delta_1\right\} \cup \left\{(q_0, a, q) \mid (q_0^2, a, q) \in \delta_2\right\}$.

Intersection:

Let $\mathcal{A}_1 = (Q_1, \Sigma, \delta_1, q_0^1, F_1)$ and $\mathcal{A}_2 = (Q_2, \Sigma, \delta_2, q_0^2, F_2)$ be two NFAs.
Then $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ defines $L(\mathcal{A}_1) \cap L(\mathcal{A}_2)$, where

- $Q = Q_1 \times Q_2$, $q_0 = (q_0^1, q_0^2)$, $F = F_1 \times F_2$,
- $((q_1, q_2), a, (q_1', q_2')) \in \delta$ iff $(q_1, a, q_1') \in \delta_1$ and $(q_2, a, q_2') \in \delta_2$.

Complementation:

Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a DFA.
Then $\mathcal{A}' = (Q, \Sigma, \delta, q_0, Q \backslash F)$ defines $\Sigma^* \backslash L(\mathcal{A})$.

Homomorphisms and inverse homomorphisms

Definition of homomorphism:

A mapping $h : \Sigma^* \to \Gamma^*$ such that $h(\varepsilon) = \varepsilon$ and $h(xy) = h(x)h(y)$.

Remark: A homomorphism is determined by $h(a)$'s for every $a \in \Sigma$.

Example: $\Sigma = \{a, b\}$ and $\Pi = \{c, d\}$, $h(a) = cc, h(b) = dd$.

Theorem. Regular languages are closed under homomorphisms and inverse homomorphisms.

Homomorphism.

Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a NFA and $h : \Sigma^* \to \Gamma^*$ be a homomorphism.
Then $\mathcal{A}' = (Q', \Gamma, \delta', q_0, F)$ defines $h(L(\mathcal{A}))$, where Q' and δ' are defined as follows,

If $(q, a, q') \in \delta$ *and* $h(a) = b_1 \dots b_k$, *then*
add $(k-1)$*-new states* $p_1, \dots, p_{k-1}$ *and*
add transitions $(q, b_1, p_1), \dots, (p_{k-1}, b_k, q')$.
In particular, if $h(a) = \varepsilon$, *then* $(q, \varepsilon, q') \in \delta'$.

Claim. $\mathcal{A}'$ defines $h(L(\mathcal{A}))$.

Homomorphisms and inverse homomorphisms

Definition of homomorphism:

A mapping $h : \Sigma^* \rightarrow \Gamma^*$ such that $h(\varepsilon) = \varepsilon$ and $h(xy) = h(x)h(y)$.

Remark: A homomorphism is determined by $h(a)$'s for every $a \in \Sigma$.

Example: $\Sigma = \{a, b\}$ and $\Pi = \{c, d\}$, $h(a) = cc, h(b) = dd$.

Theorem. Regular languages are closed under homomorphisms and inverse homomorphisms.

Inverse homomorphism.

Let $\mathcal{A} = (Q, \Gamma, \delta, q_0, F)$ be a NFA and $h : \Sigma^* \rightarrow \Gamma^*$ be a homomorphism. Then $\mathcal{A}' = (Q, \Sigma, \delta', q_0, F)$ defines $h^{-1}(L(\mathcal{A}))$, where δ' is defined as follows,

$$(q, a, q') \in \delta' \text{ iff } (q, h(a), q') \in \delta^*.$$

Outline

Nonemptiness

Nonemptiness problem:

Given a NFA $\mathcal{A}$, is $L(\mathcal{A})$ nonempty?

Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$.
Compute inductively the set of states R reachable from q_0 as follows.

- $R_0 = \{q_0\}$,
- $R_{i+1} = R_i \cup \{q \mid \exists p \in R_i, a \in \Sigma.(p, a, q) \in \delta\}$.

$L(\mathcal{A})$ is nonempty iff $R \cap F \neq \emptyset$.

Language inclusion

Language inclusion problem:

Given a NFA $\mathcal{A}$ and a NFA $\mathcal{B}$, is $L(\mathcal{A}) \subseteq L(\mathcal{B})$?

Universality problem: Given a NFA $\mathcal{A}$, is $L(\mathcal{A}) = \Sigma^*$?

Theorem. Universality problem is PSPACE-complete.

Upper bound:

Suppose $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a NFA.

A natural idea:

Construct the DFA $\mathcal{A}' = (Q', \Sigma, \delta', q'_0, F')$ by subset construction, decide the emptiness of $\mathcal{A}'' = (Q', \Sigma, \delta', q'_0, Q' \backslash F')$.

Nevertheless,

$\mathcal{A}'$ has exponentially many states.

Results from complexity theory.

- Reachability in directed graphs can be solved in NLOGSPACE

Language inclusion

Language inclusion problem:

Given a NFA $\mathcal{A}$ and a NFA $\mathcal{B}$, is $L(\mathcal{A}) \subseteq L(\mathcal{B})$?

Universality problem: Given a NFA $\mathcal{A}$, is $L(\mathcal{A}) = \Sigma^*$?

Theorem. Universality problem is PSPACE-complete.

Upper bound:

Suppose $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a NFA.

A natural idea:

Construct the DFA $\mathcal{A}' = (Q', \Sigma, \delta', q_0', F')$ by subset construction, decide the emptiness of $\mathcal{A}'' = (Q', \Sigma, \delta', q_0', Q' \backslash F')$.

Nevertheless,

$\mathcal{A}'$ has exponentially many states.

Results from complexity theory.

- Reachability in directed graphs can be solved in NLOGSPACE
  Universality problem can be solved in NPSPACE.
- NPSPACE = PSPACE.

Language inclusion: continued

Theorem. Universality problem is PSPACE-complete.

Lower bound: Reduction from the membership problem of PSPACE TMs.

Let $M = (Q, \Sigma, \Gamma, \delta, q_0, B, F)$ be a nondeterministic TM with space bounded by cn (c is some constant) and $w = a_1 \dots a_n$.

Successful computation of M over w:

A word $C_0\$C_1\$\dots\$C_n$ such that

- $C_0 = q_0 w B^{(c-1)n}$,
- for every $i : 1 \leqslant i \leqslant n$, $C_i \in \alpha q \beta$ with $q \in Q$ and $\alpha\beta \in \Gamma^{cn}$,
- for every $i : 0 \leqslant i < n$, $C_i \vdash_M C_{i+1}$,
- $C_n \in \Gamma^* F \Gamma^*$.

Unsuccessful computations:

Words in $(\Gamma \cup Q \cup \{\$\})^*$ that are not successful computations of M over w.

Language inclusion: continued

Let $M = (Q, \Sigma, \Gamma, \delta, q_0, B, F)$ be a nondeterministic TM with space bounded by cn (c is some constant) and $w = a_1 \dots a_n$.

Construct in PTIME a regular expression (or NFA) $r_{M,w}$ from M and w to describe the unsuccessful computations of M over w.

Let $\Sigma_C = Q \cup \Gamma \cup \$$.

A word $u \in r_{M,w}$ iff one of the following conditions holds.

1. u is not of the form $C_0\$C_1\$\dots\$C_k$ where $\forall i : 0 \leqslant i \leqslant k$, $C_i \in \Gamma^* Q \Gamma^*$,
2. u does not start with $q_0 w B^{(c-1)n}$,
3. there is j such that $C_j \not\vdash_M C_{j+1}$,
4. $C_n \notin \Gamma^* F \Gamma^*$.

Language inclusion: continued

Let $M = (Q, \Sigma, \Gamma, \delta, q_0, B, F)$ be a nondeterministic TM with space bounded by cn (c is some constant) and $w = a_1 \dots a_n$.

Construct in PTIME a regular expression (or NFA) $r_{M,w}$ from M and w to describe the unsuccessful computations of M over w.

Let $\Sigma_C = Q \cup \Gamma \cup \$$.

A word $u \in r_{M,w}$ iff one of the following conditions holds.

1. u is not of the form $C_0\$C_1\$\dots\$C_k$ where $\forall i : 0 \leqslant i \leqslant k$, $C_i \in \Gamma^* Q \Gamma^*$,

$$r_1 = \begin{array}{c}\Gamma^*\$\Sigma_C^* \cup \Gamma^* Q \Gamma^* Q \Gamma^* \$ \Sigma_C^* \cup \Sigma_C^* \$ \Gamma^* \cup \Sigma_C^* \$ \Gamma^* Q \Gamma^* Q \Gamma^* \cup \\ \Sigma_C^* \$ \Gamma^* \$ \Sigma_C^* \cup \Sigma_C^* \$ \Gamma^* Q \Gamma^* Q \Gamma^* \$ \Sigma_C^*\end{array}$$

2. u does not start with $q_0 w B^{(c-1)n}$,

$$r_2 = (\Sigma_C \backslash \{q_0\}) \Sigma_C^* \cup \bigcup_{i=1}^{n} \Sigma_C^i (\Sigma_C \backslash \{a_i\}) \Sigma_C^* \cup \bigcup_{i=n+1}^{cn} \Sigma_C^i (\Sigma_C \backslash \{B\}) \Sigma_C^*$$

3. there is j such that $C_j \not\vdash_M C_{j+1}$,

$$r_3 = \bigcup_{i=1}^{cn-3} \bigcup_{(\sigma_1,\sigma_2,\sigma_3,\sigma'_1,\sigma'_2,\sigma'_3) \notin f_M} \Sigma_C^* \$ \Sigma_C^i \sigma_1 \sigma_2 \sigma_3 \Sigma_C^{cn-(i+3)} \$ \Sigma_C^i \sigma'_1 \sigma'_2 \sigma'_3 \Sigma_C^*$$

4. $C_n \notin \Gamma^* F \Gamma^*$.

$$r_4 = \Sigma_C^* \$ \Gamma^* (Q \backslash F) \Gamma^*$$

Language inclusion: continued

Let $M = (Q, \Sigma, \Gamma, \delta, q_0, B, F)$ be a nondeterministic TM with space bounded by cn (c is some constant) and $w = a_1 \dots a_n$.

Construct in PTIME a regular expression (or NFA) $r_{M,w}$ from M and w to describe the unsuccessful computations of M over w.

Let $\Sigma_C = Q \cup \Gamma \cup \$$.

A word $u \in r_{M,w}$ iff one of the following conditions holds.

1. u is not of the form $C_0 \$ C_1 \$ \dots \$ C_k$ where $\forall i : 0 \leqslant i \leqslant k$, $C_i \in \Gamma^* Q \Gamma^*$,
   $r_1 = \dots$
2. u does not start with $q_0 w B^{(c-1)n}$,
   $r_2 = \dots$
3. there is j such that $C_j \not\vdash_M C_{j+1}$,
   $r_3 = \dots$
4. $C_n \notin \Gamma^* F \Gamma^*$.
   $r_4 = \dots$

Let $r_{M,w} = r_1 \cup r_2 \cup r_3 \cup r_4$.

Claim. M does not accept w iff $L(r_{M,w}) = \Sigma_C^*$.

Summary

1. NFA, right linear grammar and regular expression
2. Pumping lemma
3. Myhill-Nerode theorem
4. DFA, subset construction and minimization
5. Closure properties
6. Decision problem

Next lecture

Visibly pushdown languages