CS310 : Automata Theory 2019

Lecture 13: DFA minimization and Myhill-Nerode theorem

Instructor: Ashutosh Gupta

IITB, India

Compile date: 2019-02-01

Topic 13.1

DFA minimization

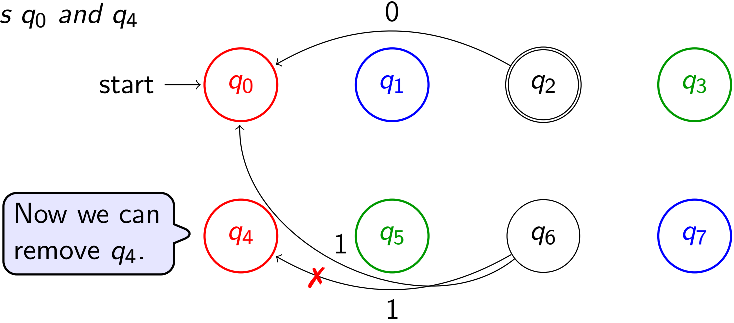
DFA minimization intuition

If states q and q' are equivalent, we can move incoming transitions for q' to q without any effect on the language recognized by the automaton.

Now we can remove q' from the automaton, which is minimization.

Example 13.1

In our running example, let us look at the incoming transitions of equivalent states q_0 and q_4



Exercise 13.1

When can we not remove q'?

DFA minimization

Let $A = (Q, \Sigma, \delta, q_0, F)$ be a DFA. Let *Blocks* be the partitions of equivalent states in A.

Let us define the minimized DFA $A' = \{Blocks, \Sigma, \delta', Block_0, Blocks_f\}$, where

- for each $B, B' \subseteq Blocks$,
  $\delta'(B, a) \triangleq B'$ if there are $q \in B$ and $q' \in B'$ such that $\delta(q, a) = q'$,
- $Blocks_f \triangleq \{B \subseteq Blocks | B \cap F \neq \emptyset\}$, and
- $q_0 \in Block_0$

Exercise 13.2

a. *Is δ' well defined?*

c. *Why $Block_0$ is unique and always exists?*

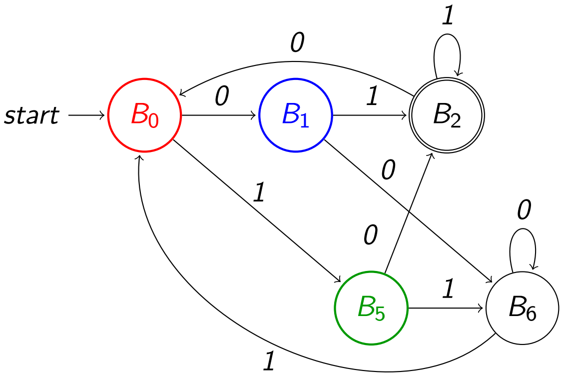
b. *Does $Blocks_f$ introduce new accepting states?*

Commentary: We are skipping the formal proof. However, the answer to the above exercises prove the correctness.

Example : DFA minimization

Example 13.2

After applying the minimization, we obtain.



$$Blocks = \{\underbrace{\{q_0, q_4\}}_{B_0}, \underbrace{\{q_1, q_7\}}_{B_1}, \underbrace{\{q_2\}}_{B_2}, \underbrace{\{q_3, q_5\}}_{B_5}, \underbrace{\{q_6\}}_{B_6}\}$$

Unique minimum DFA

Theorem 13.1

Let A be a minimized DFA. No DFA smaller than A recognizes $L(A)$.

Proof.

Assume a DFA A' such that A' has fewer states than A and $L(A) = L(A')$.
Therefore, initial states q_0 and q_0' of A and A' respectively are equivalent.

claim: All states of A are equivalent to some state of A'.

We know all states of A are reachable from its initial state(why?).
Let q be a state in A.
Let word w take A to q.
Let word w take A' to some state q'.
q and q' must be equivalent. Otherwise, q_0 and q_0' are not equivalent.(why?)

Unique minimization DFA

Proof(Contd.)

Due to the pigeonhole principle, there are states q_1 and q_2 of A such that they are equivalent to the same state of A'.

Therefore, q_1 and q_2 are equivalent.

Since A is minimized, no two states of A are equivalent. Contradiction. □

The proof makes even stronger claim. The minimized DFA is minimum up to renaming of states.

Exercise 13.3

The minimization method is also correct for NFAs. But, the uniqueness is not guaranteed. What part of the above proof does not work for NFAs?

Topic 13.2

Residual languages

Residual language

Definition 13.1

Given a language $L \subseteq \Sigma^*$ and $w \in \Sigma^*$, the residual of L with respect to w is the language

$$L^w = \{u \in \Sigma^* | wu \in L\}.$$

A language $L' \subseteq \Sigma^*$ is a residual of L if $L' = L^w$ for at least one $w \in \Sigma^*$.

Example 13.3

Let $\Sigma = \{a, b\}$ and $L = \{a, ab, ba, aab\}$

- $L^\epsilon = L$
- $L^a = \{\epsilon, b, ab\}$
- $L^{aa} = \{b\}$
- $L^{ab} = \{\epsilon\}$

Exercise 13.4

Continue considering the same language

- $L^b =$
- $L^{bb} =$
- $L^{ba} =$
- $L^{aab} =$

Distinct residual languages

For a language L, residual languages are defined with respect to each $w \in \Sigma^*$.

There will be infinitely many residual languages.

However, there may be $w, w' \in \Sigma^*$ such that $L^w = L^{w'}$.

Example 13.4

Consider language $L = aa^*b$.

$$L^a = a^*b = L^{aa} = L^{aaa}$$

Finite residual languages

There may be far fewer distinct residual languages.

May be only finitely many!!

Example 13.5

Consider language $L = aa^*b$.

- $L^a = a^*b = L^{aa} = L^{aaa}$
- $L^b = \emptyset = L^{ba}$
- $L^{ab} = \epsilon = L^{aab} = L^{aab}$
- $L^\epsilon = aa^*b$

There are no other residual languages of L.

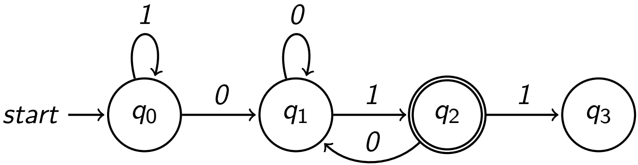
Language of a state of DFA

Definition 13.2

Let $A = (Q, \Sigma, \delta, q_0, F)$ be a DFA and $q \in Q$. The language recognized by q, denoted $L(A_q)$, is the language recognized by $A_q = (Q, \Sigma, \delta, q, F)$.

Example 13.6

Consider the following DFA A.



- $L(A_{q_1}) = 0^*1(00^*1)^*$

DFA vs residual languages

Theorem 13.2

Let $A = (Q, \Sigma, \delta, q_0, F)$ be a DFA recognizing L.

1. *For each $w \in \Sigma^*$, there is a $q \in Q$ such that $L(A_q) = L^w$.*
2. *For each $q \in Q$ reachable from q_0, there is a $w \in \Sigma^*$ such that $L(A_q) = L^w$.*

Proof.

Let $q = \hat{\delta}(q_0, w)$ in both the parts.

Clearly, $L(A_q) = L^w$. □

Exercise 13.5

In the part 2, can we construct a counterexample if q is not reachable from q_0?

Canonical deterministic automaton

For each language, we can define a deterministic automaton.

All languages have a canonical deterministic automaton

Definition 13.3

Let $L \subseteq \Sigma^*$ be a language. The canonical deterministic automaton for L is $C_L = (Q_L, \Sigma, \delta_L, L, F_L)$, where:

- *Q_L is the set of residuals of L, i.e., $Q_L = \{L_w | w \in \Sigma^*\}$,*
- *$\delta_L(K, a) = K^a$ for every $K \in Q_L$ and $a \in \Sigma$, and*
- *$F_L = \{K \in Q_L | \epsilon \in K\}$.*

Q_L may be infinite

Exercise 13.6

a. Is Q_L countable?

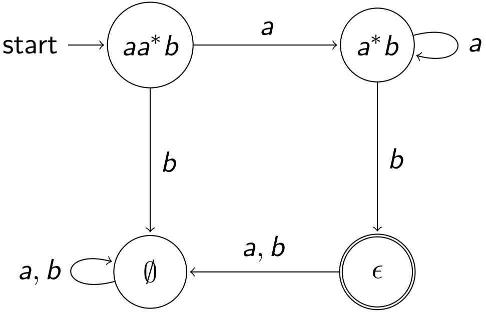
b. How many languages are there?

Example: canonical deterministic automaton

Example 13.7

Consider language $L = aa^*b$. The following are residual languages of L.

- $L^a = a^*b$
- $L^b = \emptyset = L^{ba} = L^{bb}$
- $L^{ab} = \epsilon$
- $L^\epsilon = aa^*b$



Canonical deterministic automaton C_L recognizes L

Theorem 13.3

For a language L, $C_L = (Q_L, \Sigma, \delta_L, L, F_L)$ recognizes L.

Proof.

Let $w \in \Sigma^*$. We prove by induction on the length of w $w \in L$ iff $w \in L(C_L)$.

base case:

$w = \epsilon : \epsilon \in L \quad \text{iff} \quad L \in F_L \quad \text{iff} \quad \epsilon \in L(C_L)$.

induction step:

Let $w = ax$.

$ax \in L$

iff $x \in L^a$ (definition of L^a)

iff $x \in L(C_{L^a})$ (induction hypothesis)

iff $ax \in L(C_L)$ (definition of δ_L).

□

C_L is the unique minimal DFA

Theorem 13.4

If L is regular, $C_L = (Q_L, \Sigma, \delta_L, L, F_L)$ is the unique minimal DFA up to isomorphism that recognizes L.

Proof.

Let $A = (Q, \Sigma, \delta, q_0, F)$ be a DFA that recognizes L.
Due to theorem 13.2, there is an onto mapping $M : Q \to Q_L$ such that $M(q) = L(A_q)$.
Therefore, $|Q| \geq |Q_L|$. Therefore, C_L is the minimal automaton for L.

Let A also be minimal.
claim: A is isomorphic to C_L

- $|Q| = |Q_L|$
- $\delta(q, a) = q'$ iff $L(A_{a'}) = L(A_q)^a$ iff $\delta_L(M(q), a) = M(q')$.(why?)
- $M(q_0) = L$
- $q \in F$ iff $\epsilon \in L(A_q)$ iff $L(A_q) \in F_L$ iff $M(q) \in F_L$. □

Myhill-Nerode theorem

Theorem 13.5

A language L is regular iff L has finitely many residuals.

Proof.

If L is not regular, there is no DFA recognizing it.

Therefore, the canonical deterministic automaton for L must be infinite.

Therefore, L has infinitely many residuals.

If L is regular.... □

Exercise 13.7

Complete the proof.

Commentary: In the standard presentation, the theorem is stated differently. But, the statements are equivalent.

Using Myhill-Nerode theorem

We show that L has infinitely many residuals if L is non-regular.

Does not have to be from L

- Choose an infinite sequence of words $\{w_1, w_2, ...\} \subseteq \Sigma^*$
- Show for each pair L^{w_i} and L^{w_j} there is a word that is in one and not in another

Using Myhill-Nerode theorem

Example 13.8

Consider $L = \{1^{p^2} | p \geq 0\}$.

- *Consider words* a^{i^2}.
- $a^{2i+1} \in L^{a^{i^2}}$.
- *For each* $j > i$, $j^2 + 2i + 1$ *is not a perfect square.*
- *Therefore,* $a^{2i+1} \notin L^{a^{j^2}}$.
- *Therefore, there are infinitely many residuals.*

End of Lecture 13