Minimization of DFAs

Suppose we have a DFA $M = (Q, \Sigma, \delta, q_0, F)$ recognizing $\mathcal{L}$.

Consider Q' s.t. $Q \subseteq Q'$.

What can we say about $M' = (Q', \Sigma, \delta', q_0, F)$, where

$$\delta'(q, a) = \begin{cases} \delta(q, a), & \text{if } q \in Q \\ q, & \text{if } q \in Q' \setminus Q \end{cases}$$

for any $a \in \Sigma$.

There are multiple DFAs recognizing the same language.

What if someone gave you a DFA and claimed that it recognized a language $\mathcal{L}$?

How would you check if this was indeed the case?

We would like a canonical DFA which recognizes any language.

Natural choice: DFA with fewest states

Problem:

Given a DFA $M = (Q, \Sigma, \delta, q_0, F)$,

construct a DFA $M' = (Q', \Sigma, \delta', q_0', F')$ s.t.

$$|Q| \geq |Q'|, \quad \text{and} \quad \mathcal{L}(M') = \mathcal{L}(M).$$

Key idea:

$\mathcal{L}_q$: set of strings which take M from q to some $f \in F$.

Suppose $\mathcal{L}_q = \mathcal{L}_{q'}$ for $q, q' \in Q$, and $q \neq q'$. What then?

If two states q and q' are such that the same set of strings causes the machine to accept from either state, they are equivalent in behaviour.

Enough to consider one representative for every such equivalence class.

Key idea:

One only needs different states to differentiate between different sets of strings that cause the machine to accept.

Consider an alphabet Σ, and any language $L \subseteq \Sigma^*$. $\sim_L \subseteq \Sigma^* \times \Sigma^*$

For $x, y \in \Sigma^*$, $x \sim_L y$ iff $\forall z \in \Sigma^*$: $xz \in L$ iff $yz \in L$.

Show that $\sim_L$ is an equivalence relation.

Suppose L is regular, and accepted by a DFA $M = (Q, \Sigma, \delta, q_0, F)$

Show that: For $x, y \in \Sigma^*$, if $\hat{\delta}(q_0, x) = \hat{\delta}(q_0, y)$, then $x \sim_L y$.

$[x]_L$: equivalence class of $x \in \Sigma^*$ under $\sim_L$.

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

Thm [Myhill-Nerode]:

① $\mathcal{L}$ is regular iff $\sim_L$ has finite index

② If $\mathcal{L}$ is regular, any DFA for $\mathcal{L}$ has at least as many states as the index of $\sim_L$.

Proof:

① If $\mathcal{L}$ is regular, it is accepted by some DFA $M = (Q, \Sigma, \delta, q_0, F)$.

For $x, y \in \Sigma^*$, if $x \not\sim_L y$, $\hat{\delta}(q_0, x) \neq \hat{\delta}(q_0, y)$.

So, $|Q| \geq$ index of $\sim_L$ $\longrightarrow$ also proves ② !

If $\sim_L$ has finite index, we build a DFA for $\mathcal{L}$ as follows.

$$M_L = (Q_L, \Sigma, \delta_L, q_0^L, F_L)$$

$$Q_L = \{ [x] \mid x \in \Sigma^* \}$$

$\delta_L([x], a) = [xa]$ for every $a \in \Sigma$ and $x \in \Sigma^*$

$q_0^L = [\varepsilon]$ $\quad F_L = \{ [x] \mid x \in L \}$

For M to be a DFA recognizing L, show the following

- Q_L is finite
- δ_L is well-defined:

  If $[x] = [y]$, then $\delta_L([x], a) = \delta_L([y], a)$.
- $x \in L$ iff $\hat{\delta}_L(q_0^L, x) \in F_L$.