Minimizing DFAs

CSC320

DFA State Minimization

- Given a DFA $M = (Q, \Sigma, \delta, q_0, F)$, for a language L, there is a procedure for constructing a *minimal* DFA with as few states as possible which is unique up to isomorphism (i.e., renumbering of the states).

- The process has two stages:
  - Get rid of inaccessible states.
  - Collapse "equivalent" states.

Collapsing states

$\delta(q,w) =$ state reached starting in state q and processing every symbol in w.

When can you collapse two states p and q?

We can't collapse p and q if $q \in F$ and $p \notin F$.

If we collapse p and q then we'd better collapse $\delta(p,a)$ and $\delta(q,a)$.

States p and q can be collapsed iff there is no string x *such that* $\delta(q,x) \in F$ *and* $\delta(p,x) \notin F$ or $\delta(q,x) \notin F$ *and* $\delta(p,x) \in F$

If p and q can be collapsed we say they are equivalent, and we write $p \equiv q$.

Otherwise we say they are distinguishable and there is some string which distinguishes p from q.

$\equiv$ is an equivalence relation

Claim: The relation $\equiv$ is an equivalence relation, that is, it is

- reflexive: $p \equiv p$ for all p.
- symmetric: if $p \equiv q$, then $q \equiv p$.
- transitive: if $p \equiv r$ and $r \equiv q$ then $p \equiv q$.

A state minimization algorithm

Input: DFA M with no inaccessible states:

PHASE 1: *Identify collapsable states*

- Write down a table of all pairs $\{p, q\}$.
- *Mark* $\{p, q\}$ if $p \in F$ and $q \notin F$ or vice versa.
- Repeat the following until no more changes occur:
  - For each unmarked pair $\{p, q\}$, check for each $a \in \Sigma$ whether $\{\delta(p, a), \delta(q, a)\}$ is marked. If so, then mark $\{p, q\}$.

- Partition the states of M into equivalence classes $[p] = \{q \mid \{p, q\} \text{ isn't marked}\}$. Let $Q' = \{[p] | p \in Q\}$. These will be the states of the minimal DFA.

Correctness of Phase 1

Lemma: If p and q are distinguishable then $\{p, q\}$ is marked.

Proof: We will show by induction on k that if p and q are distinguishable by a string of length k, then $\{p, q\}$ is marked.

- If $k = 0$ then exactly one of p, q is in F. and $\{p, q\}$ is marked on line 2.
- Assume the statement is true for k. Let w be a string of length $k + 1$ that distinguishes p and q. Say $w = ax$. Let $r = \delta(p, a)$ and $s = \delta(q, a)$. Then x distinguishes r from s.
- Now since $|x| = k$, $\{r, s\}$ will be marked by the algorithm (induction hypothesis), and so $\{p, q\}$ will be marked by the algorithm on line 3.

Create the DFA $\boldsymbol{M'}$

Let $M\ =\ (Q^{'}, \Sigma, \delta^{'}, [q_0], F'),$ where

- $Q^{'}\ = \{[p]\mid p \in\ Q\}$
- $\delta([p], a)\ =\ [\delta(p, a)]$
- $F^{'}\ = \{[p]|p \in\ F\}$

Correctness of the construction

Claim: M' is well-defined and each state in M' is a maximal set of states of M which are pairwise not distinguishable.

Proof: By the Lemma and the construction, each state in M' corresponds to a maximal set of states which are pairwise not distinguishable. Since this relationship is an equivalence relationship, Q' is a partition of the states of Q.

Also it must be the case that for every $q \in [p]$, and every $a \in \Sigma$ if $\delta(q, a) = r$ then $r \in [\delta(p, a)]$. Assume not, then there is a string w which distinguishes r from $\delta(p, a)$. But then aw distinguishes q from p, contradicting our assumption that $q \in [p]$.

Equivalence

Theorem: $L(M') = L(M)$

Proof: Suppose a string $w \in L(M)$. Then $\delta(q_0, w) \in F$. Since every state in $[q_0]$ is indistinguishable from q_0, $\delta'([q0], w) \in F'$.

Now suppose a string $w \in L(M')$. Then $\delta'([q0], w) \in F'$. Since every state in $[q_0]$ is indistinguishable from q_0, $\delta(q_0, w) \in F$.