Informatics 1 - Computation & Logic: Tutorial 7 Solutions

Computation: Subset Construction

Week 9: 16–20 November 2015

NFA to DFA

Recall the following definition:

A finite state machine $\mathfrak{A}$ *is an ordered 5-tuple* $\mathfrak{A} = \langle Q, \Sigma, S, A, \delta \rangle$*, where:*

- Q *is a finite set of states;*
- Σ *is an alphabet of input symbols;*
- $S \subseteq Q$ *is the set of initial, or starting states;*
- $A \subseteq Q$ *is the set of final, or accepting states; and*
- $\delta \subseteq Q \times \Sigma \times Q$*, is a set of transitions.*

We write $s \rightarrow^a t$ to mean that $(s, a, t) \in \delta$.

A finite state machine $\langle Q, \Sigma, s_0, A, \delta \rangle$ is **deterministic** if and only if it has a single initial state, s_0, and for every $q \in Q$ and every $\sigma \in \Sigma$, there is eactly one transition $\langle q, \sigma, q' \rangle \in \delta$. In this case the transition relation is the graph of a *next-state function* $N : Q \times \Sigma \rightarrow Q$.

Every FSM can be converted into a deterministic FSM that accepts exactly the same set of strings, by means of an algorithm known as the *subset construction.*

The formal definition states that the result of applying the subset construction to FSM $\langle Q, \Sigma, S, F, \delta \rangle$ is the deterministic FSM $\langle Q', \Sigma, S, F', \delta' \rangle$, where:

- Q' is a set of subsets of Q i.e. superstates
- F' is the set of all and only superstates in Q' which contain at least one state in F
- δ' is the set of derived transitions — given the original FSM $\langle Q, \Sigma, S, F, \delta \rangle$, in the new, derived machine, there is a transition from superstate A by means of symbol $\sigma \in \Sigma$ to the following superstate:

  $$\{q \mid \text{for some } q' \in A, \langle q', \sigma, q \rangle \in \delta\}$$

  i.e. the set of all states in the original machine that can be reached from *some* state in A by means of a transition labelled σ.

Take for example the following FSM, which accepts all and only the strings over $\{b, c\}$ that begin with b and end with c:

$$\mathcal{M}_1 = \langle \{0, 1, 2\}, \{b, c\}, \{0\}, \{2\}, \{\langle 0, b, 1\rangle, \langle 1, b, 1\rangle, \langle 1, c, 1\rangle, \langle 1, c, 2\rangle\}\rangle$$

It may help to draw out this FSM as a diagram. This machine is non-deterministic because there are two distinct transitions from state 1 for the symbol c.

Let us now consider how to convert $\mathcal{M}_1$ into a deterministic FSM, $\mathcal{M}_2$, that accepts exactly the same set of strings.

First of all, note that the set of states in $\mathcal{M}_1$ is $\{0, 1, 2\}$. The states in the deterministic machine $\mathcal{M}_2$ will be *subsets* of those states, which we call *superstates*.

Step 1: Identify initial superstate

The first step in converting $\mathcal{M}_1$ into a deterministic machine is to identify the initial superstate of the new machine. To do this we just take the initial states of $\mathcal{M}_1$, i.e. $\{0\}$, and view it as a superstate. Thus, the new machine, $\mathcal{M}_2$, starts off as follows:

$$\mathcal{M}_2 = \langle \{\{0\}\}, \{b, c\}, \{0\}, \emptyset, \emptyset\rangle$$

Again, it might be useful to draw this partial machine. Note that as of yet the machine $\mathcal{M}_2$ does not contain any accept states or transitions. It has just the single state, $\{0\}$, that is also the initial state.

Step 2: Incrementally add transitions

The next part of the subset construction involves incrementally building up the partial FSM until every superstate has exactly one transition for each symbol in the alphabet $\{b, c\}$. First, we identify those transitions that are missing — the partial machine has one state $\{0\}$ and this state lacks two transitions: one for symbol b and one for symbol c. Thus we must find:

- a transition from superstate $\{0\}$ for symbol b, and,
- a transition from superstate $\{0\}$ for symbol c

Given some non-deterministic FSM $\langle Q, \Sigma, s_0, F, \delta\rangle$, a superstate $A \in \wp(Q)$ and a symbol $\sigma \in \Sigma$, the superstate in the new machine that is reached from superstate A by means of symbol σ is defined as follows:

$$\{q \mid \text{for some } q' \in A, \langle q', \sigma, q\rangle \in \delta\}$$

Thus:

- the transition from superstate $\{0\}$ for symbol b is:

$$\{q \mid \text{for some } q' \in \{0\}, \langle q', b, q\rangle \in \{\langle 0, b, 1\rangle, \langle 1, b, 1\rangle, \langle 1, c, 1\rangle, \langle 1, c, 2\rangle\}\} = \{1\}$$

- the transition from superstate $\{0\}$ for symbol c is:

$$\{q \mid \text{for some } q' \in \{0\}, \langle q', c, q\rangle \in \{\langle 0, b, 1\rangle, \langle 1, b, 1\rangle, \langle 1, c, 1\rangle, \langle 1, c, 2\rangle\}\} = \emptyset$$

Thus, we have two new transitions to add to our partial machine: $\langle\{0\}, b, \{1\}\rangle$ and $\langle\{0\}, c, \emptyset\rangle$. Adding the new transitions and the new superstates $\{1\}$ and $\emptyset$ to $\mathcal{M}_2$, the partial machine is now:

$$\mathcal{M}_2 = \langle\{\{0\}, \{1\}, \emptyset\},\ \{b, c\},\ \{0\},\ \emptyset,\ \{\langle\{0\}, b, \{1\}\rangle,\ \langle\{0\}, c, \emptyset\rangle\}\rangle$$

This partial machine still has transitions missing, so we continue the construction:

- the transition from superstate $\{1\}$ for symbol b is:

$$\{q \mid \text{for some } q' \in \{1\}, \langle q', b, q\rangle \in \{\langle 0, b, 1\rangle, \langle 1, b, 1\rangle, \langle 1, c, 1\rangle, \langle 1, c, 2\rangle\}\} = \{1\}$$

- the transition from superstate $\{1\}$ for symbol c is:

$$\{q \mid \text{for some } q' \in \{1\}, \langle q', c, q\rangle \in \{\langle 0, b, 1\rangle, \langle 1, b, 1\rangle, \langle 1, c, 1\rangle, \langle 1, c, 2\rangle\}\} = \{1, 2\}$$

- the transition from superstate $\emptyset$ for symbol b is:

$$\{q \mid \text{for some } q' \in \emptyset, \langle q', b, q\rangle \in \{\langle 0, b, 1\rangle, \langle 1, b, 1\rangle, \langle 1, c, 1\rangle, \langle 1, c, 2\rangle\}\} = \emptyset$$

- the transition from superstate $\emptyset$ for symbol c is:

$$\{q \mid \text{for some } q' \in \emptyset, \langle q', c, q\rangle \in \{\langle 0, b, 1\rangle, \langle 1, b, 1\rangle, \langle 1, c, 1\rangle, \langle 1, c, 2\rangle\}\} = \emptyset$$

Thus the superstate $\{1, 2\}$ is added to machine $\mathcal{M}_2$, along with four new transitions: $\langle\{1\}, b, \{1\}\rangle$, $\langle\{1\}, c, \{1, 2\}\rangle$, $\langle\emptyset, b, \emptyset\rangle$, and $\langle\emptyset, c, \emptyset\rangle$. The partial machine is now (again, it may be useful to draw out the partial machine):

$$\mathcal{M}_2 = \langle\{\{0\}, \{1\}, \emptyset, \{1, 2\}\},\ \{b, c\},\ \{0\},\ \emptyset,\ \delta\rangle$$

where $\delta = \{\langle\{0\}, b, \{1\}\rangle,\ \langle\{0\}, c, \emptyset\rangle,\ \langle\{1\}, b, \{1\}\rangle,\ \langle\{1\}, c, \{1, 2\}\rangle,\ \langle\emptyset, b, \emptyset\rangle,\ \langle\emptyset, c, \emptyset\rangle\}$

There is *still* one superstate $\{1, 2\}$ which lacks the appropriate transitions, so we continue:

- the transition from superstate $\{1, 2\}$ for symbol b is:

$$\{q \mid \text{for some } q' \in \{1, 2\}, \langle q', b, q\rangle \in \{\langle 0, b, 1\rangle, \langle 1, b, 1\rangle, \langle 1, c, 1\rangle, \langle 1, c, 2\rangle\}\} = \{1\}$$

- the transition from superstate $\{1, 2\}$ for symbol c is:

$$\{q \mid \text{for some } q' \in \{1, 2\}, \langle q', c, q\rangle \in \{\langle 0, b, 1\rangle, \langle 1, b, 1\rangle, \langle 1, c, 1\rangle, \langle 1, c, 2\rangle\}\} = \{1, 2\}$$

So we add a further two transitions to $\mathcal{M}_2$: $\langle\{1,2\}, b, \{1\}\rangle$ and $\langle\{1,2\}, c, \{1,2\}\rangle$. The machine is now:

$$\mathcal{M}_2 = \langle\{\{0\}, \{1\}, \emptyset, \{1,2\}\},\ \{b,c\},\ \{0\},\ \emptyset,\ \delta_2\rangle$$

where

$$\begin{aligned}\delta_2 = \quad & \{\langle\{0\}, b, \{1\}\rangle,\ \langle\{0\}, c, \emptyset\rangle,\ \langle\{1\}, b, \{1\}\rangle,\ \langle\{1\}, c, \{1,2\}\rangle,\\ & \langle\emptyset, b, \emptyset\rangle,\ \langle\emptyset, c, \emptyset\rangle,\ \langle\{1,2\}, b, \{1\}\rangle,\ \langle\{1,2\}, c, \{1,2\}\rangle\}\end{aligned}$$

The transition set of the new machine is now complete, since every superstate has exactly one transition for each symbol in the alphabet $\{b, c\}$.

Step 3: Identify accepting superstates

The only thing that remains to do is to identify the set of accepting states in the new, deterministic FSM. Basically, *every* superstate in the new FSM that contains *any* of the accepting states in the original non-deterministic FSM are accepting states of the new, derived machine. The set of accepting states in $\mathcal{M}_1$ is $\{2\}$, so the only accepting state in $\mathcal{M}_2$ is the superstate $\{1, 2\}$.

Step 4: Full definition of deterministic FSM

In conclusion, the result of applying the subset construction to the non-deterministic FSM $\mathcal{M}_1$ is the following *deterministic* machine $\mathcal{M}_2$:

$$\mathcal{M}_2 = \langle\{\{0\}, \{1\}, \emptyset, \{1,2\}\},\ \{b,c\},\ \{0\},\ \{\{1,2\}\},\ \delta_2\rangle$$

where

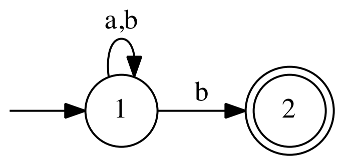
$$\begin{aligned}\delta_2 = \quad & \{\langle\{0\}, b, \{1\}\rangle,\ \langle\{0\}, c, \emptyset\rangle,\ \langle\{1\}, b, \{1\}\rangle,\ \langle\{1\}, c, \{1,2\}\rangle,\\ & \langle\emptyset, b, \emptyset\rangle,\ \langle\emptyset, c, \emptyset\rangle,\ \langle\{1,2\}, b, \{1\}\rangle,\ \langle\{1,2\}, c, \{1,2\}\rangle\}\end{aligned}$$

1. Using the subset construction, convert the following FSM:

$$\mathcal{M}_1 = \langle\{1,2\},\ \{a,b\},\ \{1\},\ \{2\},\ \{\langle 1, a, 1\rangle, \langle 1, b, 1\rangle, \langle 1, b, 2\rangle\}\rangle$$

to a deterministic FSM, $\mathcal{M}_2$, that accepts the same language.

(a) Draw the $\mathcal{M}_1$ machine:



(b) Give the complete set of states, Q_2, of $\mathcal{M}_2$:

$\{\{1\}, \{1, 2\}\}$

(c) What is the starting state, s_0, of $\mathcal{M}_2$?

$\{1\}$

(d) Give the set of transitions, δ_2, of $\mathcal{M}_2$:

$\{\langle\{1\}, a, \{1\}\rangle, \langle\{1\}, b, \{1, 2\}\rangle, \langle\{1, 2\}, a, \{1\}\rangle, \langle\{1, 2\}, b, \{1, 2\}\rangle\}$

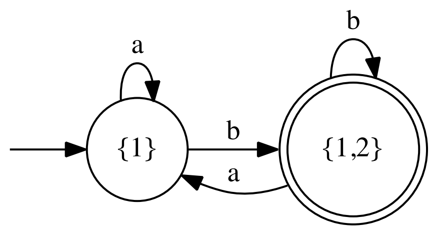
(e) What is the set of accepting states, F_2, of $\mathcal{M}_2$?

$\{\{1, 2\}\}$

(f) Now give the full definition of $\mathcal{M}_2$:

$\mathcal{M}_2 = \langle\{\{1\}, \{1, 2\}\}, \{a, b\}, \{1\}, \{\{1, 2\}\}, \delta_2\rangle$
$\delta_2 = \{\langle\{1\}, a, \{1\}\rangle, \langle\{1\}, b, \{1, 2\}\rangle, \langle\{1, 2\}, a, \{1\}\rangle, \langle\{1, 2\}, b, \{1, 2\}\rangle\}$

(g) Finally, draw the $\mathcal{M}_2$ machine:

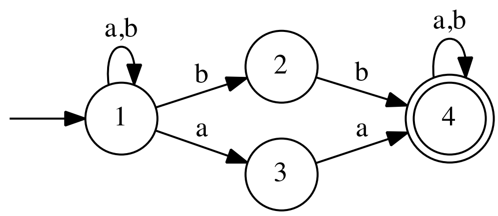


2. Convert the following non-deterministic FSM:

$$\mathcal{M}_3 = \langle \{1, 2, 3, 4\}, \{a, b\},\ \{1\},\ \{4\}, \delta_3 \rangle$$

$\delta_3 = \{\langle 1, a, 1\rangle, \langle 1, b, 1\rangle, \langle 1, b, 2\rangle, \langle 1, a, 3\rangle, \langle 2, b, 4\rangle, \langle 3, a, 4\rangle, \langle 4, a, 4\rangle, \langle 4, b, 4\rangle\}\rangle$ to a deterministic FSM, $\mathcal{M}_4$, that accepts the same language.

(a) Draw the $\mathcal{M}_3$ machine:



(b) What is the starting state, s_0, of $\mathcal{M}_4$?

$\{1\}$

(c) Give the set of transitions, δ_4, of $\mathcal{M}_4$:

$\{\langle\{1\}, a, \{1, 3\}\rangle, \langle\{1\}, b, \{1, 2\}\rangle, \langle\{1, 3\}, a, \{1, 3, 4\}\rangle, \langle\{1, 3\}, b, \{1, 2\}\rangle,$
$\langle\{1, 2\}, a, \{1, 3\}\rangle, \langle\{1, 2\}, b, \{1, 2, 4\}\rangle, \langle\{1, 3, 4\}, a, \{1, 3, 4\}\rangle,$
$\langle\{1, 3, 4\}, b, \{1, 2, 4\}\rangle, \langle\{1, 2, 4\}, a, \{1, 3, 4\}\rangle, \langle\{1, 2, 4\}, b, \{1, 2, 4\}\rangle\}$

(d) Give the set of states, Q_4, of $\mathcal{M}_4$:

$\{\{1\}, \{1, 2\}, \{1, 3\}, \{1, 2, 4\}, \{1, 3, 4\}\}$

(e) What is the set of accepting states, F_4, of $\mathcal{M}_4$?

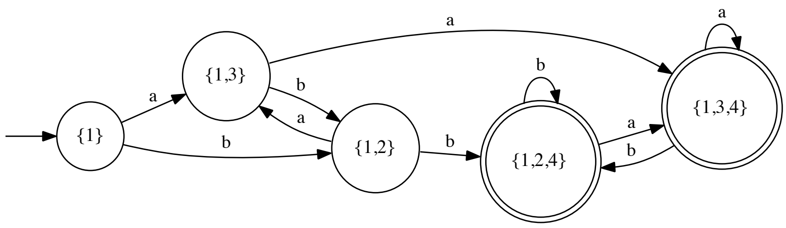
$\{\{1, 2, 4\}, \{1, 3, 4\}\}$

(f) Now give the full definition of $\mathcal{M}_4$:

$\mathcal{M}_4 = \langle\{\{1\}, \{1, 2\}, \{1, 3\}, \{1, 2, 4\}, \{1, 3, 4\}\}, \{a, b\}, \{1\}, \{\{1, 2, 4\}, \{1, 3, 4\}\}, \delta_4\rangle$

$\delta_4 = \{\langle\{1\}, a, \{1, 3\}\rangle, \langle\{1\}, b, \{1, 2\}\rangle, \langle\{1, 3\}, a, \{1, 3, 4\}\rangle, \langle\{1, 3\}, b, \{1, 2\}\rangle,$
$\langle\{1, 2\}, a, \{1, 3\}\rangle, \langle\{1, 2\}, b, \{1, 2, 4\}\rangle, \langle\{1, 3, 4\}, a, \{1, 3, 4\}\rangle,$
$\langle\{1, 3, 4\}, b, \{1, 2, 4\}\rangle, \langle\{1, 2, 4\}, a, \{1, 3, 4\}\rangle, \langle\{1, 2, 4\}, b, \{1, 2, 4\}\rangle\}$

(g) Finally, draw the $\mathcal{M}_4$ machine:



3. In the definition of an FSM we choose to write $s \to^a t$ when $\delta(s, a, t)$, but what if we chose to go the other way?

If $\mathfrak{A} = \langle Q, \Sigma, S, A, \delta \rangle$ is a FSM, we define the opposite machine, $\mathfrak{A}^{\mathrm{op}}$, to be the FSM $\mathfrak{A}^{\mathrm{op}} = \langle Q, \Sigma, A, S, \delta^{\mathrm{op}} \rangle$, where $\delta^{\mathrm{op}}(t, a, s)$ iff $\delta(s, a, t)$. Note that as well as reversing the direction of each arrow, we have swopped the roles of S and A – the initial states become final, and vice-versa.

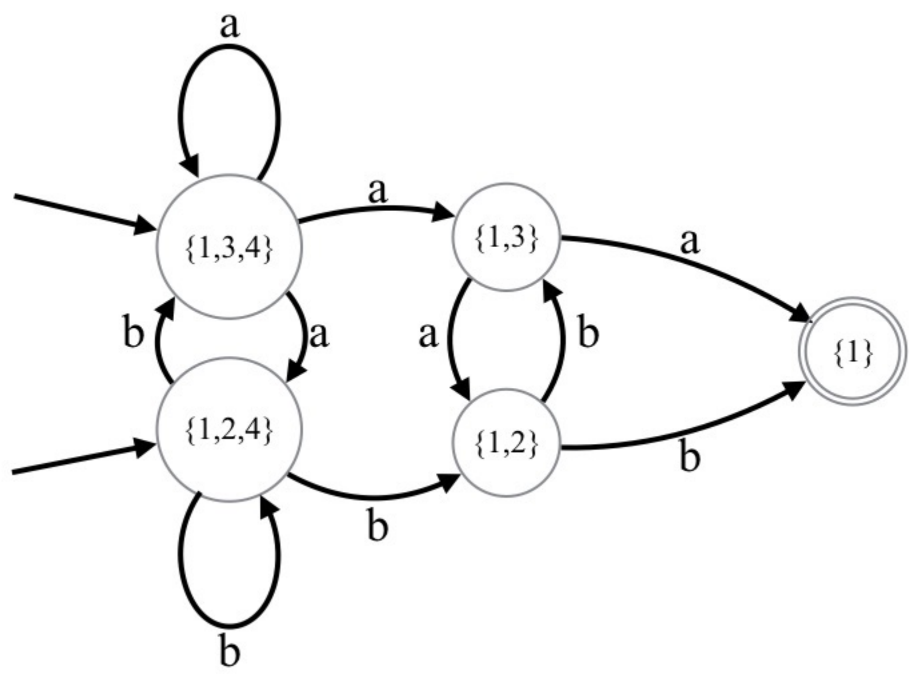
(a) What is the opposite machine of the machine $\mathcal{M}_3$ of question 2?

It is isomorphic to the original, which means that it is the same, except for a remnaming of states. To determine whether a string contains two 'a's or two'b's, it doesn't matter in which direction we read the string.

(b) If a machine, $\mathfrak{A}$, recognises the language, A, what is the language recognised by $\mathfrak{A}^{\mathrm{op}}$?

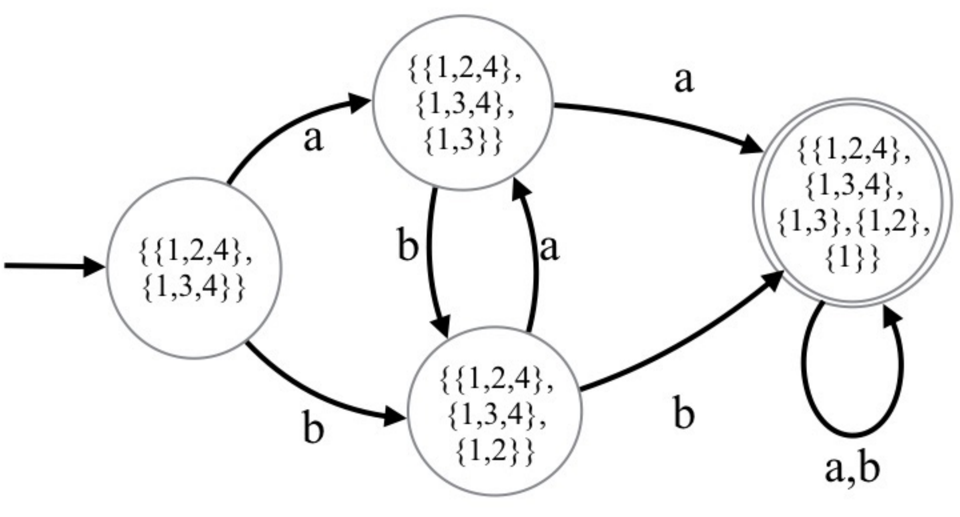
Assuming A^{op} is the language accepted by $\mathfrak{A}^{\mathrm{op}}$, then for any pair of strings s and s^{op} where s^{op} is s reversed, iff $s \in A$, $s^{\mathrm{op}} \in A^{\mathrm{op}}$.

(c) Let $\mathcal{M}_5 = \mathcal{M}_4^{\text{op}}$ be the opposite machine of the machine $\mathcal{M}_4$ of question 2. Draw $\mathcal{M}_5$.



(d) Is $\mathcal{M}_5$ deterministic? If not (strong hint) use the subset construction to compute a DFA, $\mathcal{M}_6$, equivalent to $\mathcal{M}_5$.

No, $\mathcal{M}_5$ is not deterministic; note that it has multiple start states, multiple transitions from $\{1,3,4\}$ and $\{1,3\}$ for a, multiple transitions from $\{1,2,4\}$ and $\{1,2\}$ for b, no productions from $\{1,3,4\}$, $\{1,3\}$, and $\{1\}$ for b, and none from $\{1,2,4\}$, $\{1,2\}$, and $\{1\}$ for a.



(e) Compare $\mathcal{M}_6$ with $\mathcal{M}_3$ and $\mathcal{M}_4$. Do they all recognise the same language?

They are all the same. As noted in 3a, to determine whether a string contains two 'a's or two'b's, it doesn't matter in which direction we read the string.

Brzozowski (1963) observed that, reversing the edges of a DFA produces a non-deterministic finite automaton (NFA) for the reversal of the original language, and converting this NFA to a DFA using the standard powerset construction (constructing only the reachable states of the converted DFA) leads to a minimal DFA for the same reversed language. Repeating this reversal operation a second time produces a minimal DFA for the original language.

This tutorial exercise sheet was written by Mark McConville, revised by Paolo Besana, and extended by Thomas French, Areti Manataki, and Michael Fourman. Send comments to `Michael.Fourman@ed.ac.uk`