How to write proofs

January 9, 2013

Problem. [HOPCROFT and ULLMAN, page - 73] Let L be a language. Define $\frac{1}{2}(L)$ to be

$$\{\ x \mid \text{ for some } y \text{ such that } |y| = |x|, xy \in L\ \}.$$

That is, $\frac{1}{2}(L)$ is the first halves of strings in L. Prove for each regular L that $\frac{1}{2}(L)$ is regular.

Solution. Since it is given that L is regular, let us assume that there exists a *deterministic finite automaton* $M = (Q, \Sigma, \delta, q_0, F)$ such that

$$L(M) = L.$$

We will also assume that $Q = \{q_0, q_1, \ldots, q_m\}$. So, $|Q| = m + 1$.

To prove that $\frac{1}{2}(L)$ is regular, we will construct a DFA $M' = (Q', \Sigma, \delta', q_0', F')$ such that $L(M') = \frac{1}{2}(L)$. The description of M' is as follows. The set of states in M' is

$$Q' = \{\ (q, S_0, S_1, \ldots, S_m) \mid q \in Q \text{ and } S_i \subseteq Q, \forall i\ \}.$$

The transition function δ' is defined as follows

$$\delta'\left((q_i, S_0, \ldots, S_m), a\right) = (q_j, S_0', \ldots, S_m'),$$

where

$$q_j = \delta\left(q_i, a\right) \tag{1}$$

$$S_i' = \bigcup_{\substack{b \in \Sigma, \\ s \in S_i}} \delta\left(s, b\right), \forall i. \tag{2}$$

The initial state of our new DFA will be

$$q_0' = \left(q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}\right).$$

As for the set of final states,

$$F' = \left\{(q_i, S_0, \ldots, S_m) \mid S_i \bigcap F \neq \phi\right\}.$$

Lemma 1. $\forall x \in \Sigma^*$, *if*

$$\begin{aligned} \hat{\delta}(q_0, x) &= q_j, \text{ and,} \\ S_i^x &= \bigcup_{\substack{y \in \Sigma^*, \\ |y|=|x|}} \hat{\delta}(q_i, y), \forall i, \end{aligned}$$

then

$$\hat{\delta}'((q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}), x) = (q_j, S_0^x, S_1^x, \ldots, S_m^x).$$

Proof. We will prove this by induction on the length of x. For the basis of induction, let $|x| = 0$, which means $x = \epsilon$. As the machine M is a DFA, we have

$$\begin{aligned} q_0 &= \hat{\delta}(q_0, \epsilon) \\ \text{and, } \forall i \quad S_i^\epsilon &= \bigcup_{\substack{y \in \Sigma^*, \\ |y|=|x|=0}} \hat{\delta}(q_i, y) \\ &= \{q_i\}. \end{aligned}$$

The second equality holds because $y = \epsilon$.

So, we should have

$$\begin{aligned} \hat{\delta}'((q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}), \epsilon) &= \left(\hat{\delta}(q_0, \epsilon), S_0^\epsilon, S_1^\epsilon, \ldots, S_m^\epsilon\right) \\ &= (q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}). \end{aligned}$$

By construction of δ', M' is a DFA. So, the above equality holds trivially. This establishes the base case of induction.

The induction hypothesis is –

For all $x \in \Sigma^*$ such that length of x is less than some constant k, if

$$\begin{aligned} q_j &= \hat{\delta}(q_0, x) \\ \textit{and,} \quad S_i^x &= \bigcup_{\substack{y \in \Sigma^*, \\ |y|=|x|}} \hat{\delta}(q_i, y), \forall i \end{aligned}$$

then

$$\hat{\delta}'((q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}), x) = (q_j, S_0^x, S_1^x, \ldots, S_m^x).$$

We would like to show that the above statement holds for strings of length k.

Consider an arbitrary string x of length k. Then, x can be written as

$$x = wa$$

where w is a string of length $k-1$ and a is a symbol in Σ. Let us look at the behaviour of M on w. Let

$$\begin{aligned} q_j &= \hat{\delta}(q_0, w) \text{ and} \\ S_i^w &= \bigcup_{\substack{y \in \Sigma^*, \\ |y|=|w|}} \hat{\delta}(q_i, y), \forall i. \end{aligned}$$

As the length of w is less than k, by induction hypothesis, the transition of M' on w is

$$\hat{\delta}'\left(\left(q_0,\{q_0\},\{q_1\},\ldots,\{q_m\}\right),w\right)=\left(q_j,S_0^w,S_1^w,\ldots,S_m^w\right).$$

Now, let us see how M behaves on x.

$$\begin{aligned}\hat{\delta}\left(q_0,x\right) &= \hat{\delta}\left(q_0,wa\right)\\ &= \delta\left(\hat{\delta}\left(q_0,w\right),a\right),\ \text{(from the definition of } \hat{\delta}\ )\\ &= \delta\left(q_j,a\right)\end{aligned}$$

The transition of M' on x is

$$\begin{aligned}\hat{\delta}'\left(\left(q_0,\{q_0\},\{q_1\},\ldots,\{q_m\}\right),x\right) &= \hat{\delta}'\left(\left(q_0,\{q_0\},\{q_1\},\ldots,\{q_m\}\right),wa\right)\\ &= \hat{\delta}'\left(q_0',wa\right)\\ &= \delta'\left(\hat{\delta}'\left(q_0',w\right),a\right)\\ &= \delta'\left(\left(q_j,S_0^w,S_1^w,\ldots,S_m^w\right),a\right),\\ &\quad \text{(from induction hypothesis)}\\ &= \left(\delta\left(q_j,a\right),\bigcup_{\substack{b\in\Sigma,\\ s\in S_0^w}}\delta\left(s,b\right),\bigcup_{\substack{b\in\Sigma,\\ s\in S_1^w}}\delta\left(s,b\right),\ldots,\bigcup_{\substack{b\in\Sigma,\\ s\in S_m^w}}\delta\left(s,b\right)\right)\\ &\quad \text{(from the definition of } \delta'\ )\\ &= \left(\delta\left(q_j,a\right),S_0^x,S_1^x,\ldots,S_m^x\right)\\ &\quad \text{(recall the length of } w \text{ and the definitions of } S_i^x\ )\\ &= \left(\hat{\delta}\left(q_0,x\right),S_0^x,S_1^x,\ldots,S_m^x\right).\end{aligned}$$

This shows that the claim holds for all x of length k. ■

Theorem 1. *If $x\in\frac{1}{2}\left(L\right)$, then $x\in L(M')$.*

Proof. Let $x\in\frac{1}{2}\left(L\right)$. This implies that $\exists y\in\Sigma^*$ such that $|y|=|x|$ and $xy\in L$. Let us assume

$$\hat{\delta}\left(q_0,x\right)=q_j.$$

Since $|y|=|x|,\hat{\delta}\left(q_j,y\right)\in S_j^x$. Thus, as $\hat{\delta}\left(q_0,xy\right)\in F$, it implies that $\hat{\delta}\left(q_j,y\right)\in F$. So, $S_j^x\bigcap F\neq\phi$. Hence, the state $(q_j,S_0^x,S_1^x,\ldots,S_m^x)$ of the DFA M' is a final state. From Lemma 1,

$$\begin{aligned}\hat{\delta}'\left(\left(q_0,\{q_0\},\{q_1\},\ldots,\{q_m\}\right),x\right) &= \left(q_j,S_0^x,S_1^x,\ldots,S_m^x\right)\\ &\in F'.\end{aligned}$$

Thus, $x\in L(M')$. ■

Lemma 2. $\forall x \in \Sigma^*$, *if*

$$\hat{\delta}'\left((q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}), x\right) = (q_j, S_0, S_1, \ldots, S_m).$$

then

$$\begin{aligned} q_j &= \hat{\delta}(q_0, x), \\ \textit{and, } S_i &= S_i^x, \forall i. \end{aligned}$$

Proof. We will prove this by induction on the length of x. For the basis of induction, let $|x| = 0$, which means $x = \epsilon$. As the machine M' is a DFA, we have

$$\hat{\delta}'\left((q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}), \epsilon\right) = (q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}).$$

So, for the machine M, we should have

$$\begin{aligned} q_0 &= \hat{\delta}(q_0, \epsilon) \\ \text{and, } \forall i \quad \{q_i\} &= S_i^\epsilon \end{aligned}$$

Now, from the definitions of S_i^ϵ, we have

$$\begin{aligned} S_i^\epsilon &= \bigcup_{\substack{y \in \Sigma^*, \\ |y| = |\epsilon|}} \hat{\delta}(q_i, y) \\ &= \{q_i\} \end{aligned}$$

So, the claim holds for strings of length 0. This establishes the base case of induction.

The induction hypothesis is –

For all $x \in \Sigma^*$ such that length of x is less than some constant k, if

$$\hat{\delta}'\left((q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}), x\right) = (q_j, S_0, S_1, \ldots, S_m).$$

then

$$\begin{aligned} q_0 &= \hat{\delta}(q_0, x) \\ \textit{and, } S_i &= S_i^x \end{aligned}$$

We would like to show that the above statement holds for strings of length k.

Consider an arbitrary string x of length k. Then, x can be written as

$$x = wa$$

where w is a string of length $k-1$ and a is a symbol in Σ. Let us look at the behaviour of M' on w.

$$\hat{\delta}'\left((q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}), w\right) = (q_j, S_0, S_1, \ldots, S_m).$$

As length of w is less than k, by induction hypothesis the transition of M on w gives us

$$\begin{aligned} q_j &= \hat{\delta}(q_0, w) \\ \text{and, } S_i &= S_i^w, \forall i. \end{aligned}$$

Now, let us see how M' behaves on x.

$$\begin{aligned} \hat{\delta}'((q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}), x) &= \hat{\delta}'((q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}), wa) \\ &= \hat{\delta}'(q_0', wa) \\ &= \delta'\left(\hat{\delta}'(q_0', w), a\right) \\ &= \delta'((q_j, S_0, S_1, \ldots, S_m), a) \\ &= \left(\delta(q_j, a), \bigcup_{\substack{b\in\Sigma,\\ s\in S_0}} \delta(s, b), \bigcup_{\substack{b\in\Sigma,\\ s\in S_1}} \delta(s, b), \ldots, \bigcup_{\substack{b\in\Sigma,\\ s\in S_m}} \delta(s, b)\right) \\ &= \left(\delta(q_j, a), \bigcup_{\substack{b\in\Sigma,\\ s\in S_0^w}} \delta(s, b), \bigcup_{\substack{b\in\Sigma,\\ s\in S_1^w}} \delta(s, b), \ldots, \bigcup_{\substack{b\in\Sigma,\\ s\in S_m^w}} \delta(s, b)\right) \\ &\quad (\text{ from induction hypothesis }) \\ &= (\delta(q_j, a), S_0^x, S_1^x, \ldots, S_m^x) \\ &= \left(\hat{\delta}(q_0, x), S_0^x, S_1^x, \ldots, S_m^x\right). \end{aligned}$$

This shows that the claim holds for all x of length k. ■

Theorem 2. *If $x \in L(M')$, then $x \in \frac{1}{2}(L)$.*

Proof. Let

$$\hat{\delta}'((q_0, \{q_0\}, \{q_1\}, \ldots, \{q_m\}), x) = (q_j, S_0, S_1, \ldots, S_m).$$

As x is accepted by $L(M')$, from the definition of F', we have

$$S_j \bigcap F \neq \phi.$$

From Lemma 2, we have

$$S_j = S_j^x.$$

Thus, there exists $y \in \Sigma^*$, such that $|y| = |x|$ and

$$\hat{\delta}(q_j, y) \in F.$$

So,

$$\begin{aligned} \hat{\delta}(q_0, xy) &= \hat{\delta}\left(\hat{\delta}(q_0, x), y\right) \\ &= \hat{\delta}(q_j, y) \\ &\in F. \end{aligned}$$

Hence, we have $x \in \frac{1}{2}(L)$. ■