COL352 Lecture 12

Contents

1 Recap

Discussion in previous class.

2 Definitions

3 Content

Theorem 1

Myhill-Nerode Theorem (Part 1)

If L is regular, then $=_L$ has finitely many equivalence classes.

Proof. L is regular, so there exists DFA D that recognizes L. So the number of states of D is at least the number of equivalence classes of D, which is at least the number of equivalence classes of $=_L$. Hence, this claim follows since there are finitely many states in D. □

Theorem 2

Myhill-Nerode Theorem (Part 2)

Suppose $L \subseteq \Sigma^*$ is a language such that $=_L$ has finitely many equivalence classes, say k. Then L is recognized by a DFA $D(L)$ with k states. (Note that this would imply that $D(L)$ would be a minimum-state DFA for L).

Proof. Consider $Q = \{w \in \Sigma^* \mid w \text{ is the lexicographically smallest string in its equivalence class of } =_L\}$.

$=_L$ has finitely many equivalence classes, so Q is finite, and $|Q| = k$.

Define $rep : \Sigma^* \to Q$ as $rep(x)$ = the lexicographically smallest string in the equivalence class of x under $=_L$.

Initial state of $D(L) = \epsilon$ (since $\epsilon \in Q$, as it is the lexicographically smallest string in its equivalence class).

Accepting states of $D(L)$, $A = \{w \in Q \mid w \in L\}$.

Transition function of $D(L) : \forall w \in Q, a \in \Sigma : \delta(w, a) = rep(wa)$.

Claim 2.1

$\hat{\delta}(\epsilon, x) = rep(x)$

Proof. By induction on $|x|$. If $|x| = 0$, then $x = \epsilon$, so since $\hat{\delta}(\epsilon, \epsilon) = \epsilon$, the base case is done.

Now suppose $x = x'a$ for $x' \in \Sigma^*, a \in \Sigma$. Then we have $\hat{\delta}(\epsilon, x) = \delta(\hat{\delta}(\epsilon, x'), a) = \delta(rep(x'), a) = rep(rep(x')a) = rep(x'a) = rep(x)$. The second last equality follows from $=_L$, since by definition, $x' =_L rep(x') \implies x'a =_L rep(x')a \implies x =_L rep(x')a \implies rep(x) = rep(rep(x')a)$. The first equality from the definition of $\hat{\delta}$, the second from the induction hypothesis and the third from the definition of δ. ☐

Claim 2.2

$D(L)$ recognizes L.

Proof. Let x be an arbitrary string in Σ^*. Then we have

$$\begin{aligned} x \in L &\iff rep(x) \in L \\ &\iff rep(x) \in A \\ &\iff \hat{\delta}(\epsilon, x) \in A \\ &\iff x \text{ is accepted by } D(L) \end{aligned}$$

The first follows using ϵ in the definition of $=_L$, the second from the definition of A and of rep, the third from the previous claim, and the last from the definition of a DFA. ☐

Observe that this proves Myhill-Nerode theorem part 2.

☐

Question 1

How does $\sim_{D(L)}$ look like?

Answer. The number of equivalence classes of $\sim_{D(L)}$ is equal to the number of equivalence classes of $=_L$. Since $D(L)$ recognizes L, $\sim_{D(L)}$ refines $=_L$.

Question 2

How do we show that all DFAs with k states that recognize L are isomorphic?

Answer. Suppose DFA D recognizes L and has exactly as many states as $D(L)$. So $\sim_D$ is the same as $\sim_{D(L)}$, which is the same as $=_L$. Each equivalence class of $\sim_D$ represents a state of D, and each equivalence class of $\sim_{D(L)}$ represents a state of $D(L)$, and this gives the required isomorphism. Fill in the details – exercise.

Question 3

Given a DFA $D = (Q, \Sigma, \delta, q_0, A)$, recognizing some language $L \subseteq \Sigma^*$, construct $D(L)$.

Answer. Idea: We know the thing about how $\sim_D$ refines $=_L$.

For $q \in Q$, define $C_q = \{x \in \Sigma^* \mid \hat{\delta}(q_0, x) = q\}$. If it isn't the null set, it is an equivalence class of $\sim_D$.

We do the following steps.

1. Remove unreachable states, i.e., q such that $C_q = \emptyset$.
2. Idea: Figure out, for each q, q' in Q, whether C_q, C_q' are contained in the same equialence class of $=_L$ and if so, merge them somehow.

   Define relation $\equiv$ on Q as $q \equiv q'$ if C_q, C_q' are contained in the same equivalence class of $=_L$. Then we would merge q, q' if $q \equiv q'$.