Lecture 16

The Myhill–Nerode Theorem

Let $R \subseteq \Sigma^*$ be a regular set. Recall from Lecture 15 that a *Myhill–Nerode relation for R* is an equivalence relation $\equiv$ on Σ^* satisfying the following three properties:

(i) $\equiv$ is a *right congruence*: for any $x, y \in \Sigma^*$ and $a \in \Sigma$,

$$x \equiv y \Rightarrow xa \equiv ya;$$

(ii) $\equiv$ *refines* R: for any $x, y \in \Sigma^*$,

$$x \equiv y \Rightarrow (x \in R \iff y \in R);$$

(iii) $\equiv$ is of *finite index*; that is, $\equiv$ has only finitely many equivalence classes.

We showed that there was a natural one-to-one correspondence (up to isomorphism of automata) between

- deterministic finite automata for R with input alphabet Σ and with no inaccessible states, and
- Myhill–Nerode relations for R on Σ^*.

This is interesting, because it says we can deal with regular sets and finite automata in terms of a few simple, purely algebraic properties.

In this lecture we will show that there exists a *coarsest* Myhill–Nerode relation $\equiv_R$ for any given regular set R; that is, one that every other Myhill–Nerode relation for R refines. The notions of *coarsest* and *refinement* will be defined below. The relation $\equiv_R$ corresponds to the unique minimal DFA for R.

Recall from Lecture 15 the two constructions

- $M \mapsto \equiv_M$, which takes an arbitrary DFA $M = (Q, \Sigma, \delta, s, F)$ with no inaccessible states accepting R and produces a Myhill–Nerode relation $\equiv_M$ for R:

$$x \equiv_M y \overset{\text{def}}{\Longleftrightarrow} \widehat{\delta}(s,x) = \widehat{\delta}(s,y);$$

- $\equiv \mapsto M_\equiv$, which takes an arbitrary Myhill–Nerode relation $\equiv$ on Σ^* for R and produces a DFA $M_\equiv = (Q, \Sigma, \delta, s, F)$ accepting R:

$$\begin{aligned}
[x] &\overset{\text{def}}{=} \{y \mid y \equiv x\}, \\
Q &\overset{\text{def}}{=} \{[x] \mid x \in \Sigma^*\}, \\
s &\overset{\text{def}}{=} [\epsilon], \\
\delta([x], a) &\overset{\text{def}}{=} [xa], \\
F &\overset{\text{def}}{=} \{[x] \mid x \in R\}.
\end{aligned}$$

We showed that these two constructions are inverses up to isomorphism.

Definition 16.1 A relation $\equiv_1$ is said to *refine* another relation $\equiv_2$ if $\equiv_1 \subseteq \equiv_2$, considered as sets of ordered pairs. In other words, $\equiv_1$ *refines* $\equiv_2$ if for all x and y, $x \equiv_1 y$ implies $x \equiv_2 y$. For equivalence relations $\equiv_1$ and $\equiv_2$, this is the same as saying that for every x, the $\equiv_1$-class of x is included in the $\equiv_2$-class of x. □

For example, the equivalence relation $x \equiv y \bmod 6$ on the integers refines the equivalence relation $x \equiv y \bmod 3$. For another example, clause (ii) of the definition of Myhill–Nerode relations says that a Myhill–Nerode relation $\equiv$ for R refines the equivalence relation with equivalence classes R and $\Sigma^* - R$.

The relation of *refinement* between equivalence relations is a partial order: it is reflexive (every relation refines itself), transitive (if $\equiv_1$ refines $\equiv_2$ and $\equiv_2$ refines $\equiv_3$, then $\equiv_1$ refines $\equiv_3$), and antisymmetric (if $\equiv_1$ refines $\equiv_2$ and $\equiv_2$ refines $\equiv_1$, then $\equiv_1$ and $\equiv_2$ are the same relation).

If $\equiv_1$ refines $\equiv_2$, then $\equiv_1$ is the *finer* and $\equiv_2$ is the *coarser* of the two relations. There is always a finest and a coarsest equivalence relation on any set U, namely the *identity relation* $\{(x,x) \mid x \in U\}$ and the *universal relation* $\{(x,y) \mid x, y \in U\}$, respectively.

Now let $R \subseteq \Sigma^*$, regular or not. We define an equivalence relation $\equiv_R$ on Σ^* in terms of R as follows:

$$x \equiv_R y \overset{\text{def}}{\iff} \forall z \in \Sigma^* \ (xz \in R \iff yz \in R). \tag{16.1}$$

In other words, two strings are equivalent under $\equiv_R$ if, whenever you append the same string to both of them, the resulting two strings are either both in R or both not in R. It is not hard to show that this is an equivalence relation for any R.

We show that for any set R, regular or not, the relation $\equiv_R$ satisfies the first two properties (i) and (ii) of Myhill–Nerode relations and is the coarsest such relation on Σ^*. In case R is regular, this relation is also of finite index, therefore a Myhill–Nerode relation for R. In fact, it is the coarsest possible Myhill–Nerode relation for R and corresponds to the unique minimal finite automaton for R.

Lemma 16.2 *Let $R \subseteq \Sigma^*$, regular or not. The relation $\equiv_R$ defined by (16.1) is a right congruence refining R and is the coarsest such relation on Σ^*.*

Proof. To show that $\equiv_R$ is a right congruence, take $z = aw$ in the definition of $\equiv_R$:

$$\begin{aligned} x \equiv_R y &\Rightarrow \forall a \in \Sigma \ \forall w \in \Sigma^* (xaw \in R \iff yaw \in R) \\ &\Rightarrow \forall a \in \Sigma \ (xa \equiv_R ya). \end{aligned}$$

To show that $\equiv_R$ refines R, take $z = \epsilon$ in the definition of $\equiv_R$:

$$x \equiv_R y \Rightarrow (x \in R \iff y \in R).$$

Moreover, $\equiv_R$ is the coarsest such relation, because any other equivalence relation $\equiv$ satisfying (i) and (ii) refines $\equiv_R$:

$$\begin{aligned} & x \equiv y \\ & \Rightarrow \forall z \ (xz \equiv yz) && \text{by induction on } |z|\text{, using property (i)} \\ & \Rightarrow \forall z \ (xz \in R \iff yz \in R) && \text{property (ii)} \\ & \Rightarrow x \equiv_R y && \text{definition of } \equiv_R. \end{aligned}$$

□

At this point all the hard work is done. We can now state and prove the *Myhill–Nerode theorem*:

Theorem 16.3 **(Myhill–Nerode theorem)** *Let $R \subseteq \Sigma^*$. The following statements are equivalent:*

(a) R is regular;

(b) there exists a Myhill–Nerode relation for R;

(c) the relation $\equiv_R$ is of finite index.

Proof. (a) $\Rightarrow$ (b) Given a DFA M for R, the construction $M \mapsto \equiv_M$ produces a Myhill–Nerode relation for R.

(b) $\Rightarrow$ (c) By Lemma 16.2, any Myhill–Nerode relation for R is of finite index and refines $\equiv_R$; therefore $\equiv_R$ is of finite index.

(c) $\Rightarrow$ (a) If $\equiv_R$ is of finite index, then it is a Myhill–Nerode relation for R, and the construction $\equiv \mapsto M_\equiv$ produces a DFA for R. □

Since $\equiv_R$ is the unique coarsest Myhill–Nerode relation for a regular set R, it corresponds to the DFA for R with the fewest states among all DFAs for R.

The collapsing algorithm of Lecture 14 actually gives this automaton. Suppose $M = (Q, \Sigma, \delta, s, F)$ is a DFA for R that is already collapsed; that is, there are no inaccessible states, and the collapsing relation

$$p \approx q \stackrel{\text{def}}{\Longleftrightarrow} \forall x \in \Sigma^* \ (\widehat{\delta}(p,x) \in F \Longleftrightarrow \widehat{\delta}(q,x) \in F)$$

is the identity relation on Q. Then the Myhill–Nerode relation $\equiv_M$ corresponding to M is exactly $\equiv_R$:

$$\begin{aligned}
& x \equiv_R y \\
\Longleftrightarrow\ & \forall z \in \Sigma^* \ (xz \in R \Longleftrightarrow yz \in R) && \text{definition of } \equiv_R \\
\Longleftrightarrow\ & \forall z \in \Sigma^* \ (\widehat{\delta}(s,xz) \in F \Longleftrightarrow \widehat{\delta}(s,yz) \in F) && \text{definition of acceptance} \\
\Longleftrightarrow\ & \forall z \in \Sigma^* \ (\widehat{\delta}(\widehat{\delta}(s,x),z) \in F \Longleftrightarrow \widehat{\delta}(\widehat{\delta}(s,y),z) \in F) \\
& && \text{Homework 1, Exercise 3} \\
\Longleftrightarrow\ & \widehat{\delta}(s,x) \approx \widehat{\delta}(s,y) && \text{definition of } \approx \\
\Longleftrightarrow\ & \widehat{\delta}(s,x) = \widehat{\delta}(s,y) && \text{since } M \text{ is collapsed} \\
\Longleftrightarrow\ & x \equiv_M y && \text{definition of } \equiv_M.
\end{aligned}$$

An Application

The Myhill–Nerode theorem can be used to determine whether a set R is regular or nonregular by determining the number of $\equiv_R$-classes. For example, consider the set

$$A = \{a^n b^n \mid n \geq 0\}.$$

If $k \neq m$, then $a^k \not\equiv_A a^m$, since $a^k b^k \in A$ but $a^m b^k \notin A$. Therefore, there are infinitely many $\equiv_A$-classes, at least one for each a^k, $k \geq 0$. By the Myhill–Nerode theorem, A is not regular.

In fact, one can show that the $\equiv_A$-classes are exactly

$$G_k = \{a^k\}, \quad k \geq 0,$$

$$H_k = \{a^{n+k}b^n \mid 1 \leq n\}, \quad k \geq 0,$$
$$E = \Sigma^* - \bigcup_{k \geq 0} G_k \cup H_k = \Sigma^* - \{a^m b^n \mid 0 \leq n \leq m\}.$$

For strings in G_k, all and only strings in $\{a^n b^{n+k} \mid n \geq 0\}$ can be appended to obtain a string in A; for strings in H_k, only the string b^k can be appended to obtain a string in A; and no string can be appended to a string in E to obtain a string in A.

We will see another application of the Myhill–Nerode theorem involving two-way finite automata in Lectures 17 and 18.

Historical Notes

Minimization of DFAs was studied by Huffman [61], Moore [90], Nerode [94], and Hopcroft [59], among others. The Myhill–Nerode theorem is due independently to Myhill [91] and Nerode [94] in slightly different forms.