CS 4510 : Automata and Complexity : Spring 2010
Myhill-Nerode Theorem

Subrahmanyam Kalyanasundaram

January 26, 2010

Myhill-Nerode theorem gives a necessary and sufficient condition for a language to be regular. The pumping lemma gives only a necessary condition for a language to be regular, and Myhill-Nerode theorem does one better in that respect. The theorem was proved by John Myhill and Anil Nerode in 1958. We state the necessary definitions and the theorem below, and explain the proof in a top-down manner.

Definition 1 (Distinguishable by a language)**.** *Let x, y be strings and L be a language over the same alphabet Σ. x, y are said to be* distinguishable by L *if $\exists z \in \Sigma^*$ such that $xz \in L$ and $yz \notin L$ or vice versa.*

Also, if x and y are not distinguishable by L, they are said to be indistinguishable by L*. This is denoted by $x \equiv_L y$.*

Exercise 1
Show that $\equiv_L$ is an equivalence relation.

Definition 2 (Pairwise Distinguishable)**.** *Let L be a language and X be a set of strings. X is* pairwise distinguishable by L *if every two **distinct** strings $x, y \in X$ are distinguishable by L.*

Definition 3 (Index)**.** *The* index *of a language L is the size of the largest set X of strings which is pairwise distinguishable by L.*

These are the definitions we need. The theorem is stated below.

Theorem 1 (Myhill-Nerode Theorem)**.** *Language L is regular if and only if it has a finite index. Moreover, the index of L is equal to the size of the smallest DFA which recognizes L.*

The proof of the theorem follows by the following two lemmas.

Lemma 1. *If L is recognized by a DFA with k states, then index$(L) \leq k$.*

Lemma 2. *If index$(L) = k < \infty$, then there exists a DFA with k states that recognizes L.*

The proof of the theorem follows very easily once we assume the above two lemmas.

Proof of Myhill-Nerode Theorem. We have by definition, L is regular $\iff$ there exists a DFA which recognizes L.

Suppose L is regular. Choose the smallest (least number of states) DFA that recognizes L. Let k be the number of states of this DFA. By lemma 1, $\text{index}(L) \leq k$. In particular

$$\text{index}(L) \leq \text{ Size of the smallest DFA that recognizes } L$$

This also implies that L has finite index as well. This shows one direction of the theorem.

For the other direction, let $\text{index}(L) = k < \infty$. Then by lemma 2 there exists a DFA with k states that recognizes L. This shows the other direction of the theorem, that L is regular. This also shows that

$$\text{Size of the smallest DFA that recognizes } L \leq \text{index}(L)$$

By the above two inequalities, we can conclude that $\text{index}(L) =$ the size of the smallest DFA that recognizes L. □

Now let us see the proofs of the lemmas. Note that we use the notation $\delta(q, x)$ for strings $x \in \Sigma^*$. $\delta(q, x)$ denotes the state that the DFA reaches starting from q after reading the string x.

Proof of Lemma 1. We prove this by contradiction. We have a language L recognized by a DFA with k states. We assume, for the sake of contradiction, that $\text{index}(L) > k$. This means that there is a set X, $|X| > k$, such that X is pairwise distinguishable by L.

Let M be the DFA with k states that recognizes L. Let q_0 be the start state of M. We have $|X| > k$. Then by pigeonhole principle, there exist two distinct $x, y \in X$ such that $\delta(q_0, x) = \delta(q_0, y)$.

Now we shall assert that x, y are pairwise indistinguishable, thus contradicting our assumption that $\text{index}(L) > k$. Consider any $z \in \Sigma^*$.

$$\delta(q_0, xz) = \delta(\delta(q_0, x), z) = \delta(\delta(q_0, y), z) = \delta(q_0, yz)$$

So $xz \in L \iff yz \in L$. x, y are pairwise indistinguishable, and hence by contradiction, $\text{index}(L) \leq k$. □

Proof of Lemma 2. Suppose $\text{index}(L) = k$. We have to show the existence of a DFA with k states that recognizes L. We shall in fact, construct such a DFA. Let $X = \{x_1, x_2, \ldots, x_k\}$ be a largest set which is pairwise distinguishable by L (the existence of X follows by the definition of index).

We build DFA $M = (Q, \Sigma, \delta, q_0, F)$. Σ is chosen to be the same as the alphabet of L. We choose $Q = \{q_1, q_2, \ldots, q_k\}$. It would be useful to think of each q_i as corresponding to the string $x_i \in X$.

For each $a \in \Sigma$, we define $\delta(q_i, a)$ as follows. Consider the string $x_i a$. We have $x_i a \equiv_L x_j$ for some $x_j \in X$. This follows from the fact that X is a largest pairwise distinguishable set by L. If $x_i \not\equiv_L x_j$ for any x_j, this would mean that $X \cup \{x_i a\}$ would be a bigger pairwise distinguishable set by L. Now define $\delta(q_i, a) = q_j$.

Similarly, we have the empty string $\varepsilon \equiv_L q_m$ for some q_m. Set $q_m = q_0$ as the starting state.

We **claim** (and shall prove soon) that $\delta(q_i, w) = q_j \iff x_i w \equiv_L x_j$.

Define $F = \{q_i \mid x_i \in L\}$. We have defined all the components of the DFA M now. All that remains is to show that this DFA recognizes L. Suppose $x \in L$. Then $x \equiv_L x_i$ for some x_i, such that $x_i \in L$

(**Why?**). By the above claim,

$$x = \varepsilon x \equiv_L x_i \Longrightarrow \delta(q_0, x) = q_i$$

But $q_i \in F$ since $x_i \in L$. So x is recognized by M.
If $x \notin L$, then $x \equiv_L x_i$ for some $x_i \notin L$. Like above, we get that $\delta(q_0, x) = q_i$. Again $q_i \notin F$ since $x_i \notin L$. So in this case M does not recognize x. So M recognizes exactly the language L. □

The only part that we are yet to prove is the claim above.

Proof of Claim. The claim is that $\delta(q_i, w) = q_j \Longleftrightarrow x_i w \equiv_L x_j$.
We prove this by induction on $|w|$, the length of w.

- When $|w| = 0$, $w = \varepsilon$. In this case $\delta(q_i, w) = \delta(q_i, \varepsilon) = q_i$. The claim is true since $x_i\varepsilon = x_i$.
- When $|w| = 1$, the claim is true by the definition of the transition function δ.
- When $|w| > 1$, we use induction. Suppose the claim holds for all w such that $|w| \leq l$. Let w be such that $|w| = l + 1$. Let $w = va$ where $|v| = l$ and $|a| = 1, a \in \Sigma$. We have

  $$\delta(q_i, w) = \delta(q_i, va) = \delta(\delta(q_i, v), a) = \delta(q_{j_1}, a) = q_{j_2}$$

  where $q_{j_1} = \delta(q_i, v)$ and $q_{j_2} = \delta(q_{j_1}, a)$. This means that $x_{j_1} \equiv_L x_i v$ and $x_{j_2} \equiv_L x_{j_2} a$ (this follows by induction hypothesis). The claim is true since

  $$x_i w = x_i va \equiv_L x_{j_1} a \equiv_L x_{j_2}$$

□