CS 4510: Automata and Complexity **Fall 2019**

Lecture 7: Myhill-Nerode Theorem

September 11, 2019

Lecturer: Santosh Vempala *Scribe: Aditi Laddha*

Question: L regular $\Leftrightarrow$ L can be pumped?

In last class, we saw that the forward implication is true. Unfortunately, the reverse direction is not true, i.e., there exist languages that can be pumped but that are not regular.

Definition 7.1 Equivalence Relations *A relation on set S is called an equivalence relation if it satisfies the following properties:*

1. *Reflexive:* $\forall x \in S, (x, x) \in R$
2. *Symmetric:* $\forall x, y \in S, (x, y) \in R \Leftrightarrow (y, x) \in R$
3. *Transitive:* $\forall x, y, z \in S, (x, y), (y, z) \in R \Rightarrow (x, z) \in R$

An equivalence relation R partitions S into equivalence classes such that if x, y belong to the same equivalence class $\Leftrightarrow$ $(x, y) \in R$.

Examples:

1. The relation $(x, y) \in R$ if x is born in the same city as y is an equivalence relation.
2. The relation $(x, y) \in R$ if x is "friend of" y is not equivalence relation because it may not be transitive.
3. $G = (V, E)$ is an undirected graph. Define relation R on V as $(a, b) \in R$ if there exists a path from a to b in G. Equivalence classes of R are the connected components of G.

Definition 7.2 R_L *For a language L, define R_L such that $(x, y) \in R_L$ if $\forall z \in \Sigma^*, xz \in L \Leftrightarrow yz \in L$.*

In other words, there is no suffix z such that $xz \in L$ and $yz \notin L$ or vice versa. There is no distinguishing extension.

Lemma 7.3 R_L *is an equivalence relation.*

Proof: R_L is symmetric and reflexive by definition. Transitivity holds because $(x, y), (y, x') \in R_L$ then for all $z \in \Sigma^*, xz \in L \Rightarrow yz \in L \Rightarrow x'z \in L$ and $xz \notin L \Rightarrow yz \notin L \Rightarrow x'z \notin L$. Thus, $(x, x') \in R_L$. ■

Consider $L = \{x \in \{0, 1\}^* :$ number of $1's$ in x is divisible by $3\}$. Then R_L has 3 equivalence classes.

Theorem 7.4 Myhill-Nerode *A language L is regular if and only if R_L has a finite number of equivalence classes.*

Proof: If L is regular then there exists a DFA D that recognises L. For any $x \in \Sigma^*$, label x with q, the state D stops in when given x as input. There are only finitely many labels as the number of states of D is finite.

For any pair of strings x, y having the same label and any $z \in \Sigma*$, the strings xz and yz end up in the same state when given as input to D. Thus $xz \in L$ if and only if $yz \in L$. So, $(x, y) \in R_L$. String having the same label belong to the same equivalence class of R_L. Thus, number of equivalence classes of R_L is less than number of states of D and hence finite.

Given that R_L has finitely many equivalence classes, we can contruct a DFA for L as follows:

- D contains one state for each equivalence class of R_L
- The initial state q_0 is the state corresponding to the equivalence class containing the empty string
- For a state q and $a \in \Sigma$, the transition function is defined as $\delta(q, a) = q'$ such that there exists x is in the equivalence class corresponding to state q and xa is in the equivalence class of q'. This is well defined and deterministic because if two strings x, y belong to the same equivalence class or $(x, y) \in R_L$ then for any $z \in \Sigma$, $(xz, yz) \in R_L$ and hence belong to the same equivalence class(q'). If not then let w be a distinguishing extension for xz, yz, then zw is a distinguishing extension for x, y contradicting $(x, y) \in R_L$.
- F is the set of all the states corresponding to equivalence classes containing strings in L.

■

Corollary 7.5 *Number of states in the smallest DFA for a regular language L = Number of equivalence classes of R_L.*

7.0.1 Examples

1. Prove that $L = \{0^i 1^i : i \in \mathbb{N}\}$ is not regular.

   Consider strings of the pattern $x = 0^i 1^j, y = 0^{i_1} 1^{j_1}$, if $i - j \neq i_1 - j_1$, then $z = 1^{i-j}$ forms distinguishing extension as $xz \in L$ and $yz \notin L$. Hence, the number of equivalence classes for R_L is at least the number of distinct value of $\#_0(x) - \#_1(x)$ for $x \in \Sigma^*$. So R_L has infinitely many equivalence classes.

2. Prove that $L = \{0^i 1^{2i} : i \in \mathbb{N}\}$ is not regular.

   Using a similar argument, the number of equivalence classes for R_L is at least the number of distinct value of $2 \cdot \#_0(x) - \#_1(x)$ for $x \in \Sigma^*$. So R_L has infinitely many equivalence classes.

3. Let $L_1 = \{x \in \{0, 1\}^*$: number of 1's in x is divisible by k. Is L regular? If so, what is the size of the smallest DFA for L?

   Using the corollary of Myhill-Nerode theorem, it can be proved that the minimal DFA has k states because R_{L_1} has k equivalence classes.

Note: Myhill-Nerode theorem is not covered in the Sipser textbook.