Myhill Nerode Theorem

Ayush Tandon (2K22/CO/133)

September 9, 2024

1 Introduction

Myhill-Nerode Theorem is a fundamental result in automata theory and formal languages. It provides a method to prove whether a language L is **regular** or not and it is also used for **minimization of states** in DFA (Deterministic Finite Automata).

2 Prerequisites

Deterministic Finite Automata

In DFA, for each input symbol, one can determine the states to which the machine will move in a finite number of steps. A DFA is represented as a five tuple $(Q, \Sigma, \delta, q_0, F)$ where:

Q is a finite set of states, Σ is a finite **alphabet**, $\delta : Q \times \Sigma \to Q$ is the **transition function**, $q_0 \in Q$ is the **initial state** and $F \subseteq Q$ is the set of **final states**.

Indistinguishability

Given a language L and x, y are strings over Σ^*, if for every string $z \in \Sigma^*$, **either** $xz, yz \in L$ **or** $xz, yz \notin L$, then x and y are said to be **indistinguishable** over the language L.

In other words, x is **equivalent** to y with respect to L, written $x \equiv_L y$, if and only if, for any $z \in \Sigma^*$,

$$xz \in L \iff yz \in L.$$

$\equiv_L$ is an equivalence relation: it is reflexive, symmetric and transitive. Since $\equiv_L$ is an equivalence relation over Σ^*, $\equiv_L$ partitions Σ^* into disjoint sets called **equivalence classes.**

3 Theorem Defintion

A language L is *regular* if and only if the number of equivalence classes of $\equiv_L$ is **finite**. If $\equiv_L$ partitions Σ^* into n equivalence classes, then a minimal DFA recognizing L has atleast n states.

L is not regular if and only if there are infinitely many strings $w_1, w_2, \ldots$ such that for all $w_i \neq w_j$, w_i and w_j are distinguishable with respect to L.

4 Check if Langauge is Regular or not

Example: L = { a$^n b^n \mid n \geq 0$}

Consider the strings a and b. We will check equivalence for this relation and find if the number of equivalence classes is finite or not. For k and j (where $j \neq k$), take $x = a^k$ and $y = a^j$.

Take $z = b^k$, and we get:

$$xz = a^k b^k \in L$$

$$yz = a^j b^k \notin L$$

We conclude that,

$$xz \in L \text{ but } yz \notin L$$

Therefore, all such pairs a^k and a^j are distinguishable and there are **infinitely many** equivalence classes. Hence, L is **not regular.**

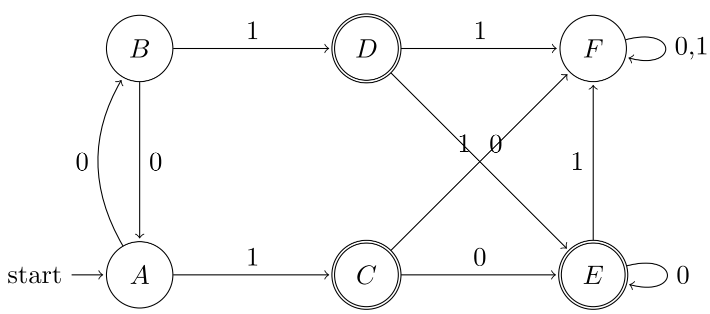
5 DFA Minimization using Myhill Nerode Theorem

5.1 Steps for minimization

1. Draw a table for all pairs of states (P, Q).
2. Mark all pairs where $P \in F$ and $Q \notin F$, where F is the set of final states.
3. If there are any unmarked pairs (P, Q) such that for all input symbols $x \in \Sigma^*$, $(\delta(P, x),\ \delta(Q, x))$ are marked, then mark (P, Q).
4. Repeat the above step until no more markings can be done.
5. Combine all **unmarked pairs** and make them a single state in the minimized DFA.

5.2 Illustration

Consider DFA:



We will also create a transition table to map all transitions.

q_0	0	1
A	B	C
B	A	D
C	E	F
D	E	F
E	E	F
F	F	F

Table 1: Transition table $\quad \delta : Q \times \Sigma \rightarrow Q$

- Set of *Final States* : {C,D,E}
- Table for all pairs: ($T[i][j]$ means i is final state and j is not a final state)
- Note: Only lower triangular matrix is drawn without diagonal values to **remove** duplicate pairs.

5.2.1 Draw State Pairs Table & Mark Final/Non-Final Pairs

	A	B	C	D	E	F
A						
B						
C	✓	✓				
D	✓	✓				
E	✓	✓				
F			✓	✓	✓	

Table 2: Pairs where $P \in F$ and $Q \notin F$

5.2.2 Mark by Transition Closure

- Unmarked Pairs are (B,A), (D,C), (E,C), (E,D), (F,A), (F,B).
- Find transitions for all states ($\delta(P, x)$, $\delta(Q, x)$) and mark where $(P, x) \in F$ and $(Q, x) \notin F$.
- Keep repeating this step until no more pairs can be marked.

Here, we found that ($\delta(F, 1)$, $\delta(A, 1)$) = (F, C), which is marked and then ($\delta(F, 0)$, $\delta(B, 0)$) = (F, A), which is also marked. So, we mark (F,A) and (F,B).
No more pairs can be marked now and so we **stop.**

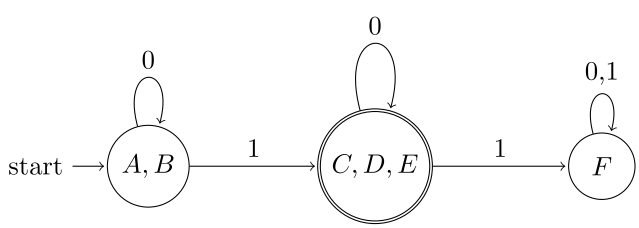
	A	B	C	D	E	F
A						
B						
C	✓	✓				
D	✓	✓				
E	✓	✓				
F	✓	✓	✓	✓	✓	

Table 3: Modified table after markings

5.2.3 Combine Unmarked States

- Unmarked Pairs are **(B,A), (D,C), (E,C) and (E,D).**
- Further, C,D,E can be combined together. So we get **3 states**, **(A,B), (C,D,E) and (F)**.
- These are 3 equivalence classes. DFA states reduced from 6 to 3.

Minimal DFA is drawn below:



6 Conclusion

The Myhill Nerode theorem is a powerful tool for classifying languages as **regular or non-regular** and minimizing DFA. It provides a systematic approach for identifying the structure of a language and optimizing the states to achieve minimal DFA.