Equivalence Relations Right Invariance Equivalence Relations Induced by DFA's The Myhill Nerode theorem Applications of the Myhill Nerode Theorem

The Myhill-Nerode Theorem

Priti Shankar

priti@csa.iisc.ernet.in

Department of Computer Science and Automation Indian Institute of Science

Outline

- Equivalence Relations
- Right Invariance
- Equivalence Relations Induced by DFA's
- The Myhill Nerode theorem
- Substitution of the Myhill Nerode Theorem

A binary relation R on a set S is a subset of $S \times S$. An equivalence relation on a set satisfies

Reflexivity: For all x in S, xRx

- Symmetry: For $x, y \in S \times Ry \iff yRx$
- Transitivity: For $x, y, z \in S$ xRy and $yRz \implies xRz$
- Every equivalence relation on *S* partitions *S* into equivalence classes. The number of equivalence classes is called the index of the relation.



A binary relation R on a set S is a subset of $S \times S$. An equivalence relation on a set satisfies

Reflexivity: For all x in S, xRxSymmetry: For $x, y \in S xRy \iff yRx$

- Transitivity: For $x, y, z \in S \times Ry$ and $yRz \implies xRz$
- Every equivalence relation on S partitions S into equivalence classes. The number of equivalence classes is called the index of the relation.



A binary relation R on a set S is a subset of $S \times S$. An equivalence relation on a set satisfies

Reflexivity: For all x in S, xRx

Symmetry: For $x, y \in S xRy \iff yRx$

Transitivity: For $x, y, z \in S xRy$ and $yRz \implies xRz$

 Every equivalence relation on S partitions S into equivalence classes. The number of equivalence classes is called the index of the relation.



A binary relation R on a set S is a subset of $S \times S$. An equivalence relation on a set satisfies

Reflexivity: For all x in S, xRx

Symmetry: For $x, y \in S xRy \iff yRx$

Transitivity: For $x, y, z \in S xRy$ and $yRz \implies xRz$

Every equivalence relation on *S* partitions *S* into equivalence classes. The number of equivalence classes is called the index of the relation.



Let
$$S = \Sigma^+$$
 where $\Sigma = \{a, b\}$.

- Define R as xRy whenever x and y both end in the same symbol of Σ .
- How many equivalence classes does R partition S into?

Let $S = \Sigma^+$ where $\Sigma = \{a, b\}$. Define R as xRy whenever x and y both end in the same symbol of Σ .

How many equivalence classes does R partition S into?

Let $S = \Sigma^+$ where $\Sigma = \{a, b\}$.

Define R as xRy whenever x and y both end in the same symbol of Σ .

How many equivalence classes does *R* partition *S* into?

An equivalence relation on Σ^* is said to be right invariant with respect to concatenation if $\forall x, y \in \Sigma^*$ and a in Σ xRy implies that xaRya.

Example:

- Let $S = \Sigma^*$ where $\Sigma = \{a, b\}$ and R be defined as follows:
- xRy if x and y have the same number of a's.
- How many equivalence classes does R partition S into?
- Is R right invariant?



An equivalence relation on Σ^* is said to be right invariant with respect to concatenation if $\forall x, y \in \Sigma^*$ and a in Σ xRy implies that xaRya.

Example:

Let $S = \Sigma^*$ where $\Sigma = \{a, b\}$ and R be defined as follows:

- *xRy* if *x* and *y* have the same number of *a's*.
- How many equivalence classes does R partition S into?
- Is R right invariant?



An equivalence relation on Σ^* is said to be right invariant with respect to concatenation if $\forall x, y \in \Sigma^*$ and a in Σ xRy implies that xaRya.

Example:

Let $S = \Sigma^*$ where $\Sigma = \{a, b\}$ and R be defined as follows:

xRy if x and y have the same number of a's.

- How many equivalence classes does R partition S into?
- Is R right invariant?



An equivalence relation on Σ^* is said to be right invariant with respect to concatenation if $\forall x, y \in \Sigma^*$ and a in Σ xRy implies that xaRya.

Example:

Let $S = \Sigma^*$ where $\Sigma = \{a, b\}$ and R be defined as follows:

xRy if *x* and *y* have the same number of *a's*. How many equivalence classes does *R* partition *S* into?

• Is R right invariant?



An equivalence relation on Σ^* is said to be right invariant with respect to concatenation if $\forall x, y \in \Sigma^*$ and a in Σ xRy implies that xaRya.

Let
$$M = (Q, \Sigma, \delta, q_0, F)$$
 be a DFA.

- Define a relation R_M as follows:
- For $x, y \in \Sigma^* x R_M y \iff \delta(q_0, x) = \delta(q_0, y)$
- Is this an equivalence relation?
- If so how many equivalence classes does it have?
 That is, what is its index?

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA. Define a relation R_M as follows:

- For $x, y \in \Sigma^* x R_M y \iff \delta(q_0, x) = \delta(q_0, y)$
- Is this an equivalence relation?
- If so how many equivalence classes does it have?
 That is, what is its index?

Let
$$M = (Q, \Sigma, \delta, q_0, F)$$
 be a DFA.
Define a relation R_M as follows:
For $x, y \in \Sigma^* x R_M y \iff \delta(q_0, x) = \delta(q_0, y)$

- Is this an equivalence relation?
- If so how many equivalence classes does it have?
 That is, what is its index?

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA.

Define a relation R_M as follows:

For
$$x, y \in \Sigma^* x R_M y \iff \delta(q_0, x) = \delta(q_0, y)$$

Is this an equivalence relation?

If so how many equivalence classes does it have?
 That is, what is its index?

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA.

Define a relation R_M as follows:

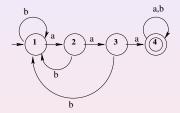
For
$$x, y \in \Sigma^* x R_M y \iff \delta(q_0, x) = \delta(q_0, y)$$

Is this an equivalence relation?

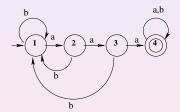
If so how many equivalence classes does it have?

That is, what is its index?





C₁: All strings not containing more than 2 consecutive *a*'s and which end in a *b*.

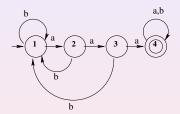


C₁: All strings not containing more than 2 consecutive *a*'s and which end in a *b*.

C₂: All strings not containing more than 2 consecutive a's and which end in a

- C₃: All strings not containing more than 2 consecutive a's and which end in aa
- C₄: All strings containing at least three consecutive a's





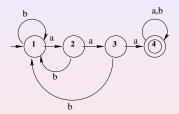
C₁: All strings not containing more than 2 consecutive *a*'s and which end in a *b*.

C₂: All strings not containing more than 2 consecutive a's and which end in a

C₃: All strings not containing more than 2 consecutive a's and which end in aa

 C₄: All strings containing at least three consecutive a's





C₁: All strings not containing more than 2 consecutive *a*'s and which end in a *b*.

C₂: All strings not containing more than 2 consecutive a's and which end in a

C₃: All strings not containing more than 2 consecutive a's and which end in aa

C₄: All strings containing at least three consecutive a's



Refinements of equivalence relations

An equivalence relation R_1 is a refinement of R_2 if $R_1 \subseteq R_2$ i.e. $(x, y) \in R_1 \implies (x, y) \in R_2$

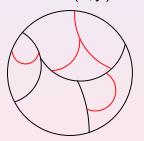


Figure: R_1 equivalence class boundaries—red lines and black lines, R_2 equivalence class boundaries—black lines





Here R_1 is not a refinement of R_2 as $(x, y) \in R_1$ does not imply $(x, y) \in R_2$



Let
$$S = \{0, 1\}^*$$

- For $x, y \in S \times R_1 y$ if x and y have the same parity.
- xR₂y if x and y have the same parity and the same length.
- What is the index of R_1 ?
- What is the index of R_2 ?
- Is one of them a refinement of the other?



```
Let S = \{0, 1\}^*
For x, y \in S \times R_1 y if x and y have the same parity.
```

- xR₂y if x and y have the same parity and the same length.
- What is the index of R_1 ?
- What is the index of R_2 ?
- Is one of them a refinement of the other?



```
Let S = \{0, 1\}^*
For x, y \in S \times R_1 y if x and y have the same parity. xR_2 y if x and y have the same parity and the same length.
```

- What is the index of R_1 ?
- What is the index of R_2 ?
- Is one of them a refinement of the other?



Let $S = \{0, 1\}^*$

For $x, y \in S \times R_1 y$ if x and y have the same parity.

 xR_2y if x and y have the same parity and the same length.

What is the index of R_1 ?

- What is the index of R_2 ?
- Is one of them a refinement of the other?



```
Let S = \{0, 1\}^*
```

For $x, y \in S \times R_1 y$ if x and y have the same parity.

 xR_2y if x and y have the same parity and the same length.

What is the index of R_1 ?

What is the index of R_2 ?

Is one of them a refinement of the other?



```
Let S = \{0, 1\}^*
```

For $x, y \in S \times R_1 y$ if x and y have the same parity.

 xR_2y if x and y have the same parity and the same length.

What is the index of R_1 ?

What is the index of R_2 ?

Is one of them a refinement of the other?



The Myhill Nerode theorem

Let $L \subseteq \Sigma_*$. The following statements are equivalent: L is recognized by a DFA

- 2 L is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.
- ⊙ Define an equivalence relation R_L as follows. For $x, y ∈ Σ^*(x, y) ∈ R_L \iff ∀z ∈ Σ^*xz ∈ L$ whenever yz ∈ L. Then R_L has finite index.

To prove this, we need to prove that

$$1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$$



The Myhill Nerode theorem

Let $L \subseteq \Sigma$ *. The following statements are equivalent:

- L is recognized by a DFA
- L is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.
- ① Define an equivalence relation R_L as follows. For $x, y \in \Sigma^* (x, y) \in R_L \iff \forall z \in \Sigma^* xz \in L$ whenever $yz \in L$. Then R_L has finite index.

To prove this, we need to prove that

$$1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$$



The Myhill Nerode theorem

Let $L \subseteq \Sigma$ *. The following statements are equivalent:

- L is recognized by a DFA
- L is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.
 - Define an equivalence relation R_L as follows. For $x, y \in \Sigma^* (x, y) \in R_L \iff \forall z \in \Sigma^* xz \in L$ whenever $yz \in L$. Then R_L has finite index.

To prove this, we need to prove that

$$1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$$



To prove $1 \implies 2$

- L is recognized by a DFA
- L is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.
 - Let $M = (Q, \Sigma, \delta, q_0, F)$ be the DFA that recognizes L.
 - The relation R_M defined earlier as follows: for $x, y \in \Sigma^* x R_M y \iff \delta(q_0, x) = \delta(q_0, y)$ is an equivalence relation which is of finite index and is also right invariant as M is a DFA.
 - L is just the union of the equivalence classes corresponding to the final states of M



To prove 1 \implies 2

- L is recognized by a DFA
- L is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.
- Let $M = (Q, \Sigma, \delta, q_0, F)$ be the DFA that recognizes L.
- The relation R_M defined earlier as follows: for $x, y \in \Sigma^* x R_M y \iff \delta(q_0, x) = \delta(q_0, y)$ is an equivalence relation which is of finite index and is also right invariant as M is a DFA.
- L is just the union of the equivalence classes corresponding to the final states of M



To prove 1 \implies 2

- L is recognized by a DFA
- L is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.

Let
$$M = (Q, \Sigma, \delta, q_0, F)$$
 be the DFA that recognizes L .

- The relation R_M defined earlier as follows: for $x, y \in \Sigma^* x R_M y \iff \delta(q_0, x) = \delta(q_0, y)$ is an equivalence relation which is of finite index and is also right invariant as M is a DFA.
- *L* is just the union of the equivalence classes corresponding to the final states of *M*



To prove $1 \implies 2$

- L is recognized by a DFA
- L is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.

Let $M=(Q,\Sigma,\delta,q_0,F)$ be the DFA that recognizes L. The relation R_M defined earlier as follows: for $x,y\in\Sigma^*$ $xR_My\iff\delta(q_0,x)=\delta(q_0,y)$ is an equivalence relation which is of finite index and is also right invariant as M is a DFA.

 L is just the union of the equivalence classes corresponding to the final states of M



To prove $1 \implies 2$

- L is recognized by a DFA
- L is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.

Let $M = (Q, \Sigma, \delta, q_0, F)$ be the DFA that recognizes L.

The relation R_M defined earlier as follows: for $x, y \in \Sigma^* x R_M y \iff \delta(q_0, x) = \delta(q_0, y)$ is an equivalence relation which is of finite index and is also right invariant as M is a DFA.

L is just the union of the equivalence classes corresponding to the final states of *M*



- 2. L is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.
- Let E be the equivalence relation in (2). We show
- We show that if $(x, y) \in E$ then $(x, y) \in R_I$, i.e.
- $(x, y) \in E \implies \forall z \in \Sigma^*, (xz, yz) \in E$ by repeated Thus R. has finite index

- 2. L is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.
- 3. Define an equivalence relation R_i as follows. For $x, y \in \Sigma^* (x, y) \in R_L \iff \forall z \in \Sigma^* xz \in L$ whenever $yz \in L$. Then R_l has finite index.
 - Let E be the equivalence relation in (2). We show
 - We show that if $(x, y) \in E$ then $(x, y) \in R_I$, i.e.
 - $(x, y) \in E \implies \forall z \in \Sigma^*, (xz, yz) \in E$ by repeated

- 2. L is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.
- 3. Define an equivalence relation R_i as follows. For $x, y \in \Sigma^* (x, y) \in R_L \iff \forall z \in \Sigma^* xz \in L \text{ whenever}$ $yz \in L$. Then R_l has finite index. Let E be the equivalence relation in (2). We show that E is a refinement of R_I . Hence since E has finite index, so does R_{l} .
 - We show that if $(x, y) \in E$ then $(x, y) \in R_I$, i.e.
 - $(x,y) \in E \implies \forall z \in \Sigma^*, (xz,yz) \in E$ by repeated

- 2. L is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.
- 3. Define an equivalence relation R_i as follows. For $x, y \in \Sigma^* (x, y) \in R_L \iff \forall z \in \Sigma^* xz \in L \text{ whenever}$ $yz \in L$. Then R_l has finite index.
 - Let E be the equivalence relation in (2). We show that E is a refinement of R_I . Hence since E has finite index, so does R_{l} .
 - We show that if $(x, y) \in E$ then $(x, y) \in R_L$, i.e. $E \subset R_l$ and hence E is a refinement of R_l .
- $(x,y) \in E \implies \forall z \in \Sigma^*, (xz,yz) \in E$ by repeated

- 2. *L* is the union of some of the equivalence classes of a right invariant equivalence relation of finite index.
- 3. Define an equivalence relation R_L as follows. For $x, y \in \Sigma^* (x, y) \in R_L \iff \forall z \in \Sigma^* xz \in L$ whenever $yz \in L$. Then R_L has finite index.

Let E be the equivalence relation in (2). We show that E is a refinement of R_L . Hence since E has finite index, so does R_L .

We show that if $(x, y) \in E$ then $(x, y) \in R_L$, i.e.

 $E \subseteq R_L$ and hence E is a refinement of R_L .

 $(x, y) \in E \implies \forall z \in \Sigma^*, (xz, yz) \in E$ by repeated use of right invariance. Hence xz is in an equivalence class in L iff yz is, thereby implying that $(x, z) \in R_L$.

Thus R. has finite index

We now prove that $3 \implies 1$

- 3. Define an equivalence relation R_L as follows. For $x, y \in \Sigma^* (x, y) \in R_L \iff \forall z \in \Sigma^*, xz \in L$ whenever $yz \in L$. Then R_L has finite index.
- 1. L is recognized by a DFA

We now prove that $3 \implies 1$

- 3. Define an equivalence relation R_L as follows. For $x, y \in \Sigma^* (x, y) \in R_L \iff \forall z \in \Sigma^*, xz \in L$ whenever $yz \in L$. Then R_L has finite index.
- 1. L is recognized by a DFA

Construct a DFA $M_L = (Q_L, \Sigma, \delta_L, q_{0L}, F_L)$.

- The state set $Q_i = \{[x] : x \in \Sigma^*\}$. The transition

- Construct a DFA $M_L = (Q_L, \Sigma, \delta_L, q_{0L}, F_L)$. We first show that R_L is right invariant. To see this assume that $(x, y) \in R_L$. Assume R_L is not right invariant. Then there exists $a \in \Sigma$ such that $(xa, ya) \notin R_L$. Let z be a "distinguishing" string for the pair (xa, ya). Then az is a distinguishing string for (x, y) contradicting the assumption that $(x, y) \in R_L$. Therefore R_L is right invariant.
- The state set $Q_L = \{[x] : x \in \Sigma^*\}$. The transition function δ_L is defined as $\delta_L([x], a) = [xa]$. The definition is consistent because had we chosen some other representative [y] of the equivalence class, [xa] = [ya] by right invariance of R_L so the transition is to the same state of M_L .

Construct a DFA $M_L = (Q_L, \Sigma, \delta_L, q_{0L}, F_L)$. We first show that R_L is right invariant. To see this assume that $(x, y) \in R_L$. Assume R_L is not right invariant. Then there exists $a \in \Sigma$ such that $(xa, ya) \notin R_L$. Let z be a "distinguishing" string for the pair (xa, ya). Then az is a distinguishing string for (x, y) contradicting the assumption that $(x, y) \in R_L$. Therefore R_L is right invariant.

Let [x] denote the equivalence class containing x. The state set $Q_L = \{[x] : x \in \Sigma^*\}$. The transition function δ_L is defined as $\delta_L([x], a) = [xa]$. The definition is consistent because had we chosen some other representative [y] of the equivalence class, [xa] = [ya] by right invariance of R_L so the transition is to the same state of M_L

Construct a DFA $M_L = (Q_L, \Sigma, \delta_L, q_{0L}, F_L)$. We first show that R_L is right invariant. To see this assume that $(x, y) \in R_L$. Assume R_L is not right invariant. Then there exists $a \in \Sigma$ such that $(xa, ya) \notin R_L$. Let z be a "distinguishing" string for the pair (xa, ya). Then az is a distinguishing string for (x, y) contradicting the assumption that $(x, y) \in R_L$. Therefore R_L is right invariant.

Let [x] denote the equivalence class containing x. The state set $Q_L = \{[x] : x \in \Sigma^*\}$. The transition function δ_L is defined as $\delta_L([x], a) = [xa]$. The definition is consistent because had we chosen some other representative [y] of the equivalence class, [xa] = [ya] by right invariance of R_L so the transition is to the same state of M_L

We need to prove that $x \in L \iff \delta_L(q_{0L}, x) \in F_L$ i.e. x is accepted by M_L

• This is easy as $\delta_L(q_{0L}, x) = [x]$ and x is accepted by M_L iff $[x] \in F_L \iff x \in L$.

We need to prove that $x \in L \iff \delta_L(q_{0L}, x) \in F_L$ i.e. x is accepted by M_L

This is easy as $\delta_L(q_{0L}, x) = [x]$ and x is accepted by M_L iff $[x] \in F_L \iff x \in L$.

- Example. Is the following language regular $L_1 = \{xy : |x| = |y|, x, y \in \Sigma^*\}$?
- Example. What about the language $L_2 = \{xy : |x| = |y| \text{ and } y \text{ ends with a 1} \}$?
- Example. What about the language $L_3 = \{xy : |x| = |y| \text{ and } y \text{ contains a 1}\}$?



Example. Is the following language regular

$$L_1 = \{xy: |x| = |y|, x, y \in \Sigma^*\} \ ?$$

- Example. What about the language $L_2 = \{xy : |x| = |y| \text{ and } y \text{ ends with a 1}\}$?
- Example. What about the language $L_3 = \{xy : |x| = |y| \text{ and } y \text{ contains a 1} \}$?



Example. Is the following language regular

$$L_1 = \{xy : |x| = |y|, x, y \in \Sigma^*\}$$
 ?

Example. What about the language

$$L_2 = \{xy : |x| = |y| \text{ and } y \text{ ends with a 1}\}$$
?

• Example. What about the language $L_3 = \{xy : |x| = |y| \text{ and } y \text{ contains a 1}\}$?



Example. Is the following language regular

$$L_1 = \{xy : |x| = |y|, x, y \in \Sigma^*\}$$
 ?

Example. What about the language

$$L_2 = \{xy : |x| = |y| \text{ and } y \text{ ends with a 1}\}$$
?

Example. What about the language

$$L_3 = \{xy : |x| = |y| \text{ and } y \text{ contains a 1}\}$$
?

Equivalence Relations Right Invariance Equivalence Relations Induced by DFA's The Myhill Nerode theorem Applications of the Myhill Nerode Theorem

For the language L_1 there are two equivalence classes of R_{L_1} . The first C_1 contains all strings of even length and the second C_2 all strings of odd length. All strings z of odd length when concatenated with any string in C_1 take us to C_2 , and all strings z of even length when concatenated with any string in C_1 take us back to C_1 which is nothing but L_1 . Similarly all strings z of even length when concatenated with any string of C_2 take us back to C_2 and all strings of odd length to C_1

Equivalence Relations Right Invariance Equivalence Relations Induced by DFA's The Myhill Nerode theorem Applications of the Myhill Nerode <u>Theorem</u>

For L_2 we have the additional constraint that y ends with a 1. Class C_2 remains the same as that for L_1 . Class C_1 is refined into classes C_1' which contains all strings of even length that end in a 1 and C_1'' which contains all strings of even length which end in a 0. Thus L_1 and L_2 are both regular.

Equivalence Relations Right Invariance Equivalence Relations Induced by DFA's The Myhill Nerode theorem Applications of the Myhill Nerode Theorem

For L_3 we have to distinguish for example, between the even length strings in the sequence 01, 0001, 000001,... and in general 0ⁱ1 for i odd, as 00 distinguishes the first string from all the others after it in the sequence (concatenation of 00 to 01 gives a string not in L_3 but concatenation of 00 to all the others gives a string in L_3), 0000 distinguishes the second from all the others after it in the sequence, and in general $(00)^{2i}$ distinguishes the i^{th} string from all the others after it in the sequence. As there are an infinite number of strings that can be distinguished from one another, the relation R_L has infinite index and therefore L_3 is not regular.