KYUSHU UNIVERSITY 100th 2011
知 の 新 世 紀 を 拓 く

形式言語理論

Formal Language Theory

7．正規言語に対する決定手続き

九州大学
KYUSHU UNIVERSITY

本日の内容

- 準備
  - 集合の対称差
- 空性判定問題
- 無限性判定問題
- 等価性判定問題
- Myhill-Nerodeの定理
- Myhill-Nerodeの定理の応用

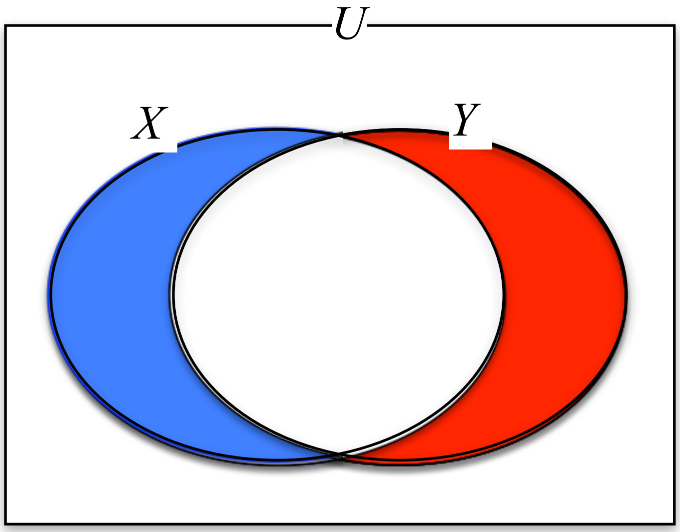
準備

集合の対称差

- 任意の集合$X,\ Y \subseteq U$に対し，

$$X \triangle Y = (X \cap (U - Y)) \cup ((U - X) \cap Y)$$

を $X,\ Y$ の対称差とよぶ.



【命題】$X = Y \Leftrightarrow X \triangle Y = \varnothing$

同値関係

【定義】 集合 A 上の関係 R について，次のように定める．

(1) R が**反射的**である
$\Leftrightarrow \forall a \in A, a \, R \, a.$

(2) R が**推移的**である
$\Leftrightarrow \forall a, b, c \in A, ((a \, R \, b \wedge b \, R \, c) \rightarrow a \, R \, c).$

(3) R が**対称的**である
$\Leftrightarrow \forall a, b \in A, (a \, R \, b \rightarrow b \, R \, a).$

【定義】 集合 A 上の関係 R が**同値関係**であるとは，
R が反射的，推移的，かつ，対称的であるときをいう．

同値関係を表す記号として，$\equiv$ がよく用いられる．

同値類

【定義】 $\equiv$ を集合 A 上の同値関係とする．任意の $x \in A$ に対して集合 $[x]_{\equiv} = \{y \mid x \equiv y\}$ を x を含む**同値類**といい，x をその代表元という．

【命題】 集合 A 上の同値関係 $\equiv$ に対し，次が成り立つ．

(1) すべての $a \in A$ に対して $a \in [a]_{\equiv}$．

(2) $A = \bigcup_{a \in A} [a]_{\equiv}$.

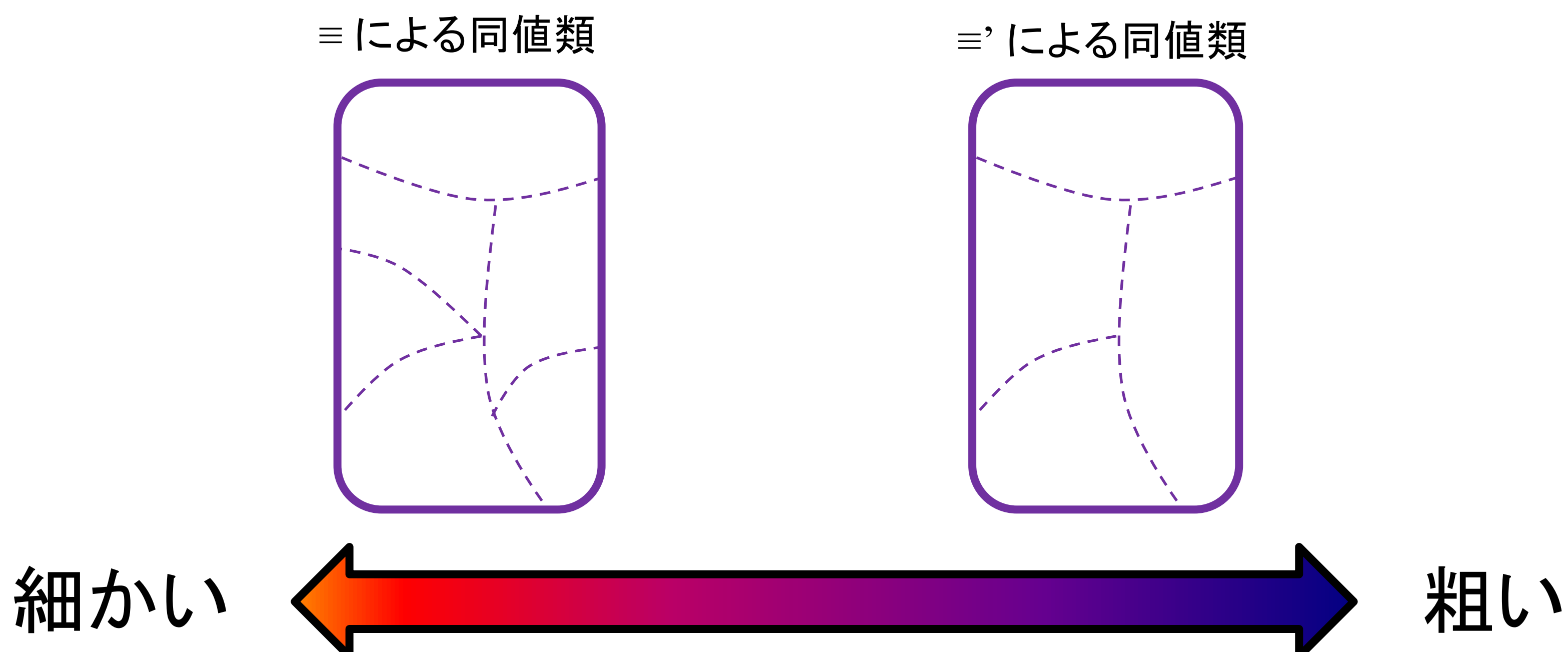
(3) すべての $a, b \in A$ に対して，$[a]_{\equiv} = [b]_{\equiv} \vee [a]_{\equiv} \cap [b]_{\equiv} = \emptyset$.

同値関係 $\equiv$ による同値類全体の集合を $A/\equiv$ で表す．
すなわち，$A/\equiv \; = \{[a]_{\equiv} \mid a \in A\}$ である．

細分

【定義】 $\equiv$, $\equiv'$ を集合 A 上の同値関係とする．すべての $a, b \in A$ に対して， $a \equiv b \Rightarrow a \equiv' b$ が成り立つとき， $\equiv$ は $\equiv'$ の細分であるという．

このとき， $\equiv'$ による同値類は $\equiv$ による同値類の和となる．



例

- 正整数 n に対し集合 $\mathbb{Z}$ 上の同値関係 $\equiv_n$ を次で定義する.
  - $a \equiv_n b \Leftrightarrow a \bmod n = b \bmod n$.
- 例えば $\equiv_6$ は $\equiv_3$ の細分である.

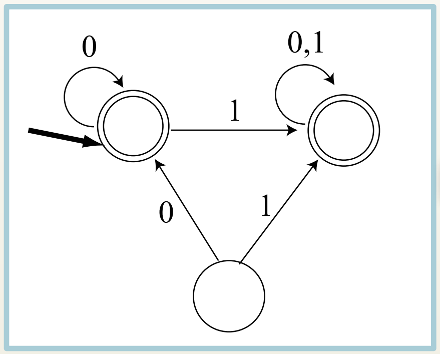
$$[0]_{\equiv_3} = [0]_{\equiv_6} \cup [3]_{\equiv_6}$$
$$[1]_{\equiv_3} = [1]_{\equiv_6} \cup [4]_{\equiv_6}$$
$$[2]_{\equiv_3} = [2]_{\equiv_6} \cup [5]_{\equiv_6}$$

空性判定問題

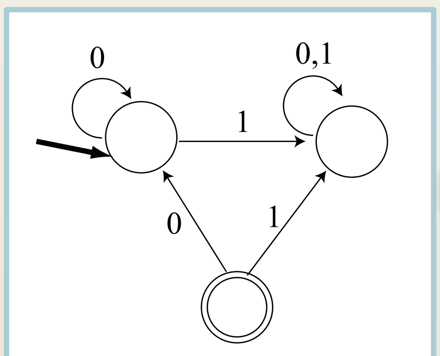
空性判定問題

DFA M → 入力 → ？ → 出力 → $L(M)=\varnothing$ ならYes そうでないならNo

例



→ ？ → No



→ ？ → Yes

空性判定問題

【定義】(空性判定問題).

入力: DFA $M=(Q, \Sigma, \delta, q_0, F)$.

出力: $L(M)=\varnothing$ ならYes，そうでないならNo.

【定理3.7】 空性判定問題を解くアルゴリズムが存在する.

定理3.7の証明

【補題3.2】 DFA M の状態数を n とするとき，次が成り立つ．

$$L(M) \neq \varnothing \Leftrightarrow \exists w \in L(M), |w| \leq n.$$

- 補題3.2に基づき，次のアルゴリズムによって空性判定問題を解くことができる．
  1. 長さが n 以下の各 $w \in \Sigma^*$ に対して：
     a. w を入力として DFA M を動作させる．
     b. M が w を受理すれば No を出力して停止．
  2. Yes を出力して停止．

QED

補題3.2の証明

- $(\Longleftarrow)$
  - 明らか.
- $(\Longrightarrow)$
  - $L(M) \neq \varnothing$ とする.
  - $L(M)$に属する最短の文字列をwとする.
  - $n \leq |w|$と仮定すると，反復補題より，$w = xyz,\ |xy| \leq n,\ |y| > 0$ となる x, y, z が存在して $xy^kz \in L(M)$ $(k=0,1,\ldots)$ となる．ここで，$k=0$とおくと，$xz \in L(M)$かつ $|xz| < |w|$ となり，w の最短性に反する.
  - よって，$|w| < n$ を得る.

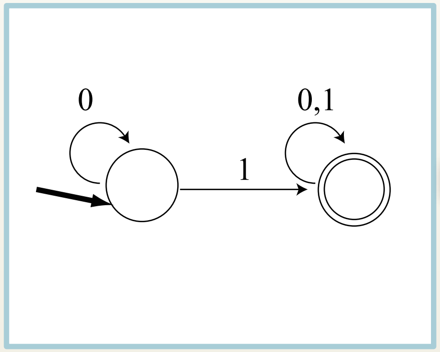
QED

無限性判定問題

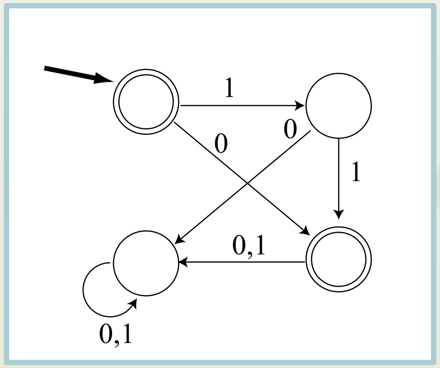
無限性判定問題

DFA M → 入力 → ？ → 出力 → $L(M)$が無限集合ならYes そうでないならNo

例



→ ？ → Yes



→ ？ → No

無限性判定問題

【定義】(無限性判定問題).
入力: DFA $M = (Q, \Sigma, \delta, q_0, F)$.
出力: $L(M)$が無限集合 ならYes, そうでないならNo.

【定理3.8】 無限性判定問題を解くアルゴリズムが存在する.

定理3.8の証明

【補題3.3】 DFA M の状態数を n とするとき，次が成り立つ.

$$|L(M)| = \infty \Leftrightarrow \exists w \in L(M),\ n \leq |w| < 2n.$$

- 補題3.3に基づき，次のアルゴリズムによって無限性判定問題を解くことができる.

1. 長さが n 以上 $2n$ 未満の各文字列 $w \in \Sigma^*$ に対して:
    a. w を入力として DFA M を動作させる.
    b. M が w を受理すれば Yes を出力して停止.
2. No を出力して停止.

QED

補題3.3の証明

- ($\Leftarrow$)
  - $n \leq |w| < 2n$ となる $w \in L(M)$ が存在すると仮定.
  - 反復補題より, $w = xyz$, $|xy| \leq n$, $|y| > 0$ となる x, y, z が存在して $xy^kz \in L(M)$ $(k=0,1,\ldots)$ となる.
  - よって, $L(M)$ は無限集合である.
- ($\Rightarrow$)
  - $L(M)$は無限集合と仮定する. $n \leq |w|$となる $w \in L(M)$ が存在する. そのような最短の w を選ぶ.
  - 反復補題より, $w = xyz$, $|xy| \leq n$, $|y| > 0$ となる x, y, z が存在して $xz \in L(M)$となる. wの選び方から$|xz|<n$となる.
  - $|y| \leq |xy| \leq n$, $|xz|<n$より$|w|<2n$である.

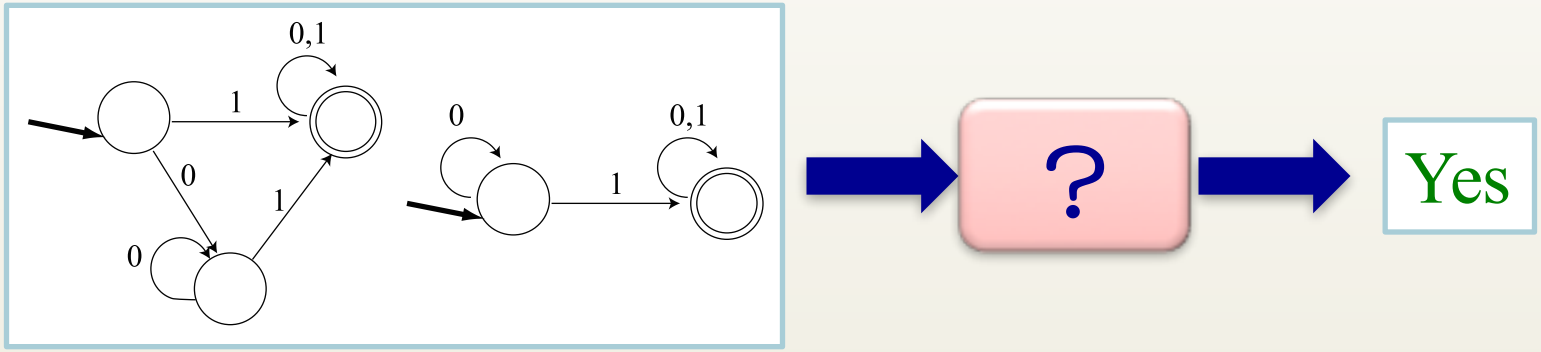
QED

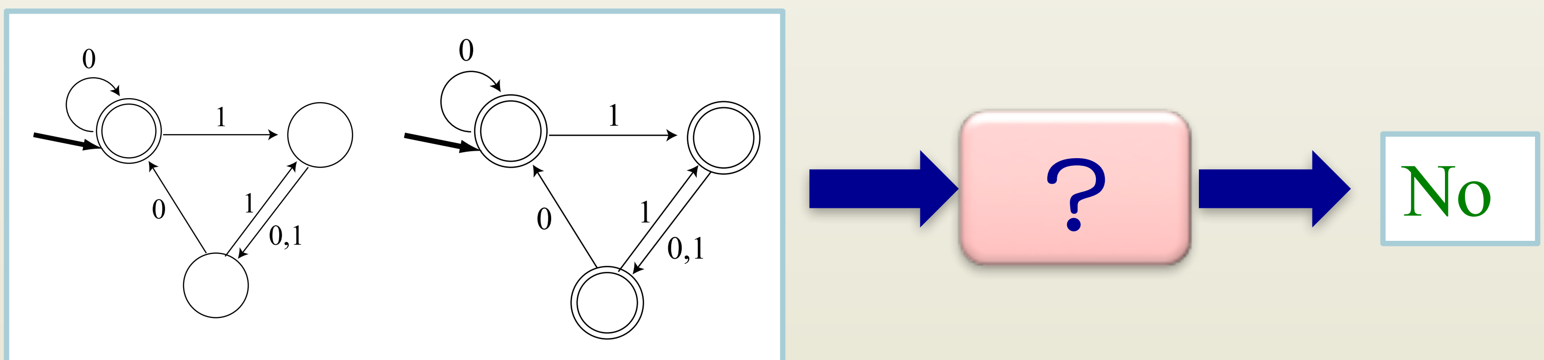
等価性判定問題

等価性判定問題

DFA M_1, M_2 → 入力 → ? → 出力 → $L(M_1)=L(M_2)$ ならYes
そうでないならNo

例





等価性判定問題

【定義】(等価性判定問題).
入力: DFA M_1, M_2.
出力: $L(M_1)=L(M_2)$ ならYes, そうでないならNo.

【定理3.9】 等価性判定問題を解くアルゴリズムが存在する.

定理3.9の証明

- 定理3.2, 3.3, 3.4より正規言語族は和集合，積集合，補集合について閉じている．したがって，
  $L(M_1)\triangle L(M_2) = (L(M_1)\cap(\Sigma^*-L(M_2)))\cup((\Sigma^*-L(M_1))\cap L(M_2))$
  は正規である．
- 正規言語 $L(M_1)\triangle L(M_2)$ を受理する DFA M_3を DFA M_1, M_2から構成できる．
- $L(M_1) = L(M_2) \Leftrightarrow L(M_1)\triangle L(M_2) = \varnothing$ が成り立つから，M_3に対して空性判定アルゴリズムを動作させることにより，M_1とM_2の等価性を判定できる．

QED

Myhill-Nerodeの定理

右不変・有限指数

【定義】 Σ^* 上の関係 R が右不変であるとは，次が成り立つときをいう.

$$x\ R\ y \Rightarrow \forall z \in \Sigma^*, xz\ R\ yz$$

【定義】 Σ^* 上の同値関係 $\equiv$ が有限指数であるとは，$\equiv$ による同値類の個数が有限であるときをいう.

DFAに基づく同値関係

【定義】 Σ 上の任意のDFA $M = (Q, \Sigma, \delta, q_0, F)$ に対して，Σ^* 上の同値関係 $\equiv^M$ を次で定義する．

$$x \equiv^M y \Leftrightarrow \delta(q_0, x) = \delta(q_0, y)$$

【補題3.4】 Σ^* 上の同値関係 $\equiv^M$ は有限指数かつ右不変である．

補題3.4の証明

- 任意の $x, y \in \Sigma^*$ に対して

$$
\begin{aligned}
x \equiv^M y &\Rightarrow \delta(q_0, x) = \delta(q_0, y) \\
&\Rightarrow \delta(\delta(q_0, x), z) = \delta(\delta(q_0, y), z) \quad (\forall z \in \Sigma^*) \\
&\Rightarrow \delta(q_0, xz) = \delta(q_0, yz) \quad (\forall z \in \Sigma^*) \\
&\Rightarrow xz \equiv^M yz \quad (\forall z \in \Sigma^*)
\end{aligned}
$$

であるから，$\equiv^M$ は右不変．

- $\equiv^M$ による同値類全体の集合 $\Sigma^*/\equiv^M$ について，$|Q| \geq |\Sigma^*/\equiv^M|$ である．よって，$\equiv^M$ は有限指数．

QED

言語に基づく同値関係

【定義】 任意の言語 $L \subseteq \Sigma^*$ に対して，Σ^* 上の同値関係 $\equiv_L$ を次で定義する.

$$x \equiv_L y$$
$$\Leftrightarrow$$
$$\forall z \in \Sigma^*,\ (xz \in L \leftrightarrow yz \in L)$$

【補題3.5】 Σ^* 上の同値関係 $\equiv_L$ は右不変である.

補題3.5の証明

- $x \equiv_L y$ とする.
- 任意の $u, v \in \Sigma^*$ に対して $z = uv$ とおくと, $\equiv_L$ の定義から, $xuv \in L \leftrightarrow yuv \in L$ となる.
- すべての $v \in \Sigma^*$ に対して $(xu)v \in L \leftrightarrow (yu)v \in L$ が成り立つので, $xu \equiv_L yu$ を得る.

QED

Myhill-Nerodeの定理

【定理3.10】 次の(1)(2)(3)は等価である.

(1) 言語 $L \subseteq \Sigma^*$ は 正規である.

(2) 有限指数で右不変な同値関係 $\equiv$ が存在し,
言語 L は $\equiv$ による同値類の和として表せる.

(3) 同値関係 $\equiv_L$ は有限指数である.

定理3.10の証明: (1)⇒(2)

- L を受理する DFA を $M=(Q,\Sigma,\delta,q_0,F)$ とし, Σ^*上の同値関係 $\equiv^M$ について考える.
- 補題3.4 より, $\equiv^M$ は有限指数で右不変である.
- L が $\equiv^M$ による同値類の和として書けることを示すためには, すべての $x\in L$ に対し $[x]_{\equiv^M}\subseteq L$ が成り立つことを示せばよい.
- $y\in[x]_{\equiv^M}$ とする. $x\equiv^M y$ より $\delta(q_0,x)=\delta(q_0,y)$. $x\in L$ であるから $\delta(q_0,x)\in F$. よって $\delta(q_0,y)\in F$, すなわち, $y\in L$ である.

QED

定理3.10の証明: (2)⇒(3)

- $x \equiv y$ とする. $\equiv$ は右不変であるから，すべての $z \in \Sigma^*$ に対して $xz \equiv yz$ が成り立つ.
- L は $\equiv$ による同値類の和で表され，xz と yz は同じ同値類に属することから，$xz \in L \Leftrightarrow yz \in L$ となる.
- すべての $z \in \Sigma^*$ に対して $xz \in L \Leftrightarrow yz \in L$ が成り立つので，$x \equiv_L y$ である.
- すべての $x, y \in \Sigma^*$ に対して，$x \equiv y \Rightarrow x \equiv_L y$ が得られた. すなわち，$\equiv$ は $\equiv_L$ の細分である.
- $\equiv$ は有限指数であるから，$\equiv_L$ も有限指数となる.

QED

定理3.10の証明: (3)⇒(1)

- 同値関係 $\equiv_L$ による同値類 $[x]_{\equiv_L}$を$[x]$と略記する.
- DFA $M_L = (Q, \Sigma, \delta, q_0, F)$ を以下のように定める.
  - $Q = \{[x] \mid x \in \Sigma^*\}$
    $\equiv_L$が有限指数なので，Qは有限集合となる.
  - $\delta([x], a) = [xa]$ $([x] \in Q, a \in \Sigma)$
  - $q_0 = [\varepsilon]$
  - $F = \{[x] \mid x \in L\}$

  δの値は，代表元の取り方によらず一意に定まる.
  すなわち，$\equiv_L$ は右不変であるので， $a \in \Sigma$に対し
  $x \equiv_L y \Rightarrow xa \equiv_L ya \Rightarrow [xa]=[ya]$.
- この M_L について，$\delta(q_0,w)=\delta([\varepsilon],w)=[\varepsilon w]=[w]$である.
- よって， $w \in L(M_L) \Leftrightarrow [w] \in F \Leftrightarrow w \in L$． すなわち，
  $L(M_L) = L$.

QED

状態数最小DFA

【定理3.11】 定理3.10の(3)⇒(1)の証明で定義したDFA M_L は, L を受理するDFA の中で状態数が最小である.

定理3.11の証明

- L を任意の正規言語とし，M を L を受理するDFAの一つとする．
- M の状態数は $|\Sigma^*/{\equiv^M}|$ 以上である.
- 同値関係 $\equiv^M$ は定理3.10 の (2) を満たす．また，$\equiv^M$ は $\equiv_L$の細分となる.
- よって，$|\Sigma^*/{\equiv^M}| \geq |\Sigma^*/{\equiv_L}|$ である.
- $\equiv_L$ に基づくDFA M_L の状態数は $|\Sigma^*/{\equiv_L}|$ に等しい.

QED

Myhill-Nerodeの定理の応用

Myhill-Nerodeの定理の応用

- Myhill-Nerodeの定理は，言語が正規でないことを証明する際に有用である.

例 $L=\{a^n b^n \mid n\geq 0\}$が正規でないことの証明

- $L=\{a^n b^n \mid n\geq 0\}$が正規であると仮定する.
- L が正規であるから，Myhill-Nerodeの定理より，同値関係 $\equiv_L$ は有限指数となる．よって，$a^i \equiv_L a^j$ となる整数 i, j $(i<j)$が存在する.
- また，$\equiv_L$ は右不変であるから， $a^i b^i \equiv_L a^j b^i$ となる.
- $a^i b^i \in L$であるから，$a^j b^i \in L$ となる．これはLの定義に反する.
- よって，背理法により言語 L は正規ではない.

QED

演習問題

- 反復補題の代わりにMyhill-Nerodeの定理を用いて以下の証明を与えよ.

【例題3.3】言語$L_3=\{w \mid w \in \{a, b\}^*, \#_a(w)=\#_b(w)\}$は正規言語でないことを証明せよ.

【例題3.4】言語$L_4=\{uu \mid u \in \{a,b\}^*\}$は正規言語でないことを証明せよ.

【演習問題3.1】言語$L=\{uu^R \mid u \in \{a,b\}^*\}$は正規言語でないことを証明せよ.