計算論A　第3回

1. 有限オートマトン
2. 正則表現と正則言語
   - 正則表現
   - 有限オートマトンから正則表現へ
3. 正則言語の性質
4. 文脈自由文法と言語
5. プッシュダウン・オートマトン
6. 文脈自由言語の性質
7. チューリングマシン

テキスト
3.1〜3.2.2節

3.1 正則表現

- 正則表現（正規表現，regular expression）
  - 言語（語の集合）を表す
  - 情報科学分野ではよく利用される表現法
    - UNIX の grep コマンド，Web ブラウザなどの検索
    - コンパイラの字句解析に用いる Lex や Flex などのコマンド
- 正則表現が表す言語＝有限オートマトンが認識する言語
  これは，後で証明する

3.1.1 正則表現の演算

- L, M：Σ 上の言語（語の集合）
- 集合和 $L \cup M$
  - $L \cup M = \{w \in \Sigma^* \mid w \in L$ または $w \in M\}$
- 連接 LM
  - $LM = \{\alpha\beta \in \Sigma^* \mid \alpha \in L,\ \beta \in M\}$
  - $L = \{001, 10, 111\}$, $M = \{\epsilon, 001\}$ のとき
    $LM = \{001, 10, 111, 001001, 10001, 111001\}$
- 閉包(スター閉包，クリーネ(Kleene)閉包) L^*
  - $L^* = \bigcup_{i \geq 0} L^i$
    - $L^0 = \{\epsilon\}$, $L^k = LL^{k-1}$　($k \geq 1$)
  - $L = \{0, 11\}$ のとき，$L^* = \{\epsilon, 0, 11, 00, 011, 110, 1111, \ldots\}$

3.1.2 正則表現の構成(1)

- アルファベット Σ 上の正則表現
  - ϵ，$\emptyset$，(，)，＋，＊ は Σ に含まれない記号
- 以下の構成法で構成される表現のみが正則表現
  $L(E)$：正則表現 E が表す言語
  - 基礎
    1. ϵ，$\emptyset$ は正則表現
       - $L(\epsilon) = \{\epsilon\}$，$L(\emptyset) = \emptyset$
    2. $a \in \Sigma$ に対し，a は正則表現
       - $L(a) = \{a\}$

3.1.2 正則表現の構成 (2)

- 再帰
  1. E，F が正則表現 $\Rightarrow$ $E+F$ も正則表現
     - $L(E+F)=L(E)\cup L(F)$
  2. E，F が正則表現 $\Rightarrow$ EF も正則表現
     - $L(EF)=L(E)L(F)$
  3. E が正則表現 $\Rightarrow$ E^* も正則表現
     - $L(E^*)=L(E)^*$
  4. E が正則表現 $\Rightarrow$ (E) も正則表現
     - $L((E))=L(E)$

例3.2 正則表現

- 正則表現：$0, 1$
  - $L(0)=\{0\}$, $L(1)=\{1\}$
- 正則表現：01
  - $L(01)=\{0\}\{1\}=\{01\}$
- 正則表現：$(01)^*$
  - $L\big((01)^*\big)=\{01\}^*=\{\epsilon, 01, 0101, 010101, 01010101, \dots\}$
- 正則表現：$(01)^*+(10)^*+0(10)^*+1(01)^*$
  - $L\big((01)^*+(10)^*+0(10)^*+1(01)^*\big)$
    $=\{w \mid w$ は 0 と 1 が交互に現れる語$\}$
- 正則表現：$(\epsilon+1)(01)^*(\epsilon+0)$
  - 上と同じ言語を表す

3.1.3 正則表現中の演算の順序

- 正則表現の括弧の省略
  - 優先順位が高い順に
    - スター演算，連接，集合和
    - $01^*+1=(0(1)^*)+1$
  - 連接は結合法則を満たす
    - $012=(01)2=0(12)$

3.2 有限オートマトンと正則表現

- 有限オートマトンが受理する言語のクラスと、正則表現で定義できる言語のクラスは一致する

(1) 任意の有限オートマトン A に対し，$L(A)$ を表す正則表現 R が存在する（3.2.1節，3.2.2節）
- A を DFA と仮定してよい

(2) 任意の正則表現 R に対し，$L(R)$ を受理する有限オートマトンが存在する（3.2.3節）
- A を ϵ-NFA と仮定してよい

3.2.1 DFA から正則表現へ (1)

【定理3.4】DFAの受理集合は正則表現で記述できる．すなわち，DFA A に対して $L = L(A)$ のとき，$L = L(R)$ を満たす正則表現 R がある．

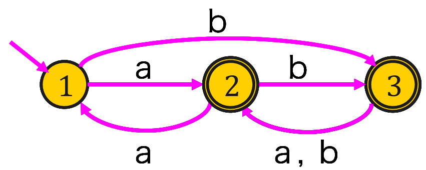
- **全頂点間の最短経路のFloydのアルゴリズムと同アイデア（動的計画法の考え方）で証明**
- **DFA** $A = (Q,\ \Sigma,\ \delta,\ q_0,\ F)$
  $Q = \{1,\ 2, \dots ,\ n\},\ q_0 = 1$ **とする**
- **正則表現** $R_{i,j}^{(k)}$ **:** $L\left(R_{i,j}^{(k)}\right) = \{\alpha = a_1 a_2 \dots a_m \mid \hat{\delta}(i,\ \alpha) = j \wedge$
  $\hat{\delta}(i,\ a_1 a_2 \dots a_h) \leq k\ \ (1 \leq h < m)\}$

状態 i から j に遷移

途中に経由する状態は$\{1,\ 2, \dots ,\ k\}$の状態のみ

3.2.1 DFA から正則表現へ (2)

- **正則表現** $R_{i,j}^{(k)}$ **:** $L\left(R_{i,j}^{(k)}\right) = \{\alpha = a_1 a_2 \dots a_m \mid \hat{\delta}(i,\ \alpha) = j \wedge$
  $\hat{\delta}(i,\ a_1 a_2 \dots a_h) \leq k\ \ (1 \leq h < m)\}$



$R_{1,3}^{(0)} = b$ **状態経由なし**

$R_{1,3}^{(1)} = b$ **状態 1 のみ経由可能**

$R_{1,3}^{(2)} = a^* b$ **状態 1，2 のみ経由可能**

$R_{1,3}^{(3)} = a^* b\left((a+b)a^* b\right)^*$ **全状態経由可能**

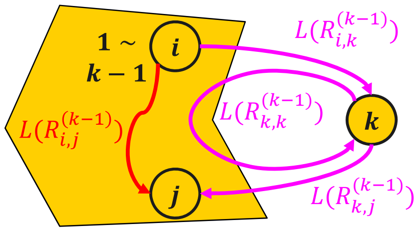
3.2.1 DFA から正則表現へ (3)

- **正則表現** $R_{i,j}^{(k)}$ **:** $L\left(R_{i,j}^{(k)}\right) = \{\alpha = a_1 a_2 \dots a_m \mid \hat{\delta}(i,\ \alpha) = j \wedge$
  $\hat{\delta}(i,\ a_1 a_2 \dots a_h) \leq k\ \ (1 \leq h < m)\}$

$L(R_{i,j}^{(0)}) = \{a \in \Sigma \mid \delta(i,\ a) = j\}\quad (i \neq j)$

$L(R_{i,i}^{(0)}) = \{a \in \Sigma \mid \delta(i,\ a) = i\} \cup \{\epsilon\}$

$L(R_{i,j}^{(k)}) = L(R_{i,j}^{(k-1)}) \cup L(R_{i,k}^{(k-1)}) \cdot L(R_{k,k}^{(k-1)})^* \cdot L(R_{k,j}^{(k-1)})$



3.2.1 DFA から正則表現へ (3)

$L(R_{i,j}^{(0)}) = \{a \in \Sigma \mid \delta(i,\ a) = j\}\quad (i \neq j)$

$L(R_{i,i}^{(0)}) = \{a \in \Sigma \mid \delta(i,\ a) = i\} \cup \{\epsilon\}$

$L(R_{i,j}^{(k)}) = L(R_{i,j}^{(k-1)}) \cup L(R_{i,k}^{(k-1)}) \cdot L(R_{k,k}^{(k-1)})^* \cdot L(R_{k,j}^{(k-1)})$

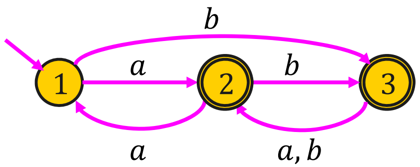
$L(R_{i,j}^{(k)})$**は，正則表現で表現できる**

- $A = (Q,\ \Sigma,\ \delta,\ q_0,\ F)\quad (Q = \{1,\ 2, \dots ,\ n\},\ q_0 = 1)$
  が認識する言語

  $L(A) = \cup_{p \in F}\, L(R_{1,p}^{(n)})$

  は，正則表現で表現できる

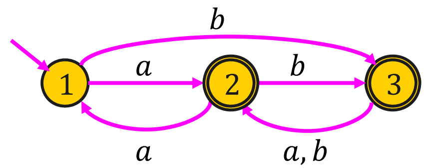
適用例 (1)



$R_{1,1}^0$: ϵ	$R_{1,2}^0$: a	$R_{1,3}^0$: b
$R_{2,1}^0$: a	$R_{2,2}^0$: ϵ	$R_{2,3}^0$: b
$R_{3,1}^0$: $\emptyset$	$R_{3,2}^0$: $a+b$	$R_{3,3}^0$: ϵ
$R_{1,1}^1$: ϵ	$R_{1,2}^1$: a	$R_{1,3}^1$: b
$R_{2,1}^1$: a	$R_{2,2}^1$: $\epsilon + aa$	$R_{2,3}^1$: $b + ab$
$R_{3,1}^1$: $\emptyset$	$R_{3,2}^1$: $a+b$	$R_{3,3}^1$: ϵ

$R_{2,3}^0 +$
$R_{2,1}^0 \cdot {R_{1,1}^0}^* \cdot R_{1,3}^0$

適用例(2)



$R_{1,1}^2$: $(aa)^*$　$R_{1,2}^2$: $a(aa)^*$　$R_{1,3}^2$: a^*b

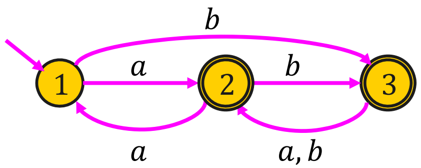
$R_{2,1}^2$: $a(aa)^*$　$R_{2,2}^2$: $(aa)^*$　$R_{2,3}^2$: a^*b

$R_{3,1}^2$: $(a+b)(aa)^*a$　$R_{3,2}^2$: $(a+b)(aa)^*$

$R_{3,3}^2$: $\epsilon + (a+b)a^*b$

$R_{3,3}^1 + R_{3,2}^1 \cdot {R_{2,2}^1}^* \cdot R_{2,3}^1$
$= \epsilon + (\mathrm{a}+\mathrm{b}) \cdot (\epsilon + aa)^* \cdot (b + ab)$
$= \epsilon + (a+b) \cdot (aa)^* \cdot (b+ab) = \epsilon + (a+b)a^*b$

適用例(3)



$R_{1,2}^3$: $R_{1,2}^2 + R_{1,3}^2 \cdot {R_{3,3}^2}^* \cdot R_{3,2}^2$

$a(aa)^* + a^*b(\epsilon + (a+b)a^*b)^*(a+b)(aa)^*$

$= a(aa)^* + a^*b\big((a+b)a^*b\big)^*(a+b)(aa)^*$

$R_{1,3}^3$: $a^*b\big((a+b)a^*b\big)^*$

オートマトンが受理する言語：

$a(aa)^* + a^*b\big((a+b)a^*b\big)^*(\epsilon + (a+b)(aa)^*)$

DFA から正則言語へ

- **全頂点間の最短経路のFloydのアルゴリズムと同アイデア（動的計画法の考え方）**
  - **手間：** $O(n^3)$　n **：オートマトンの状態数**
  - **NFA，** ϵ**-NFA にも適用可能**
- **正則表現の方程式（3.2.2節の代わりに）**
  状態遷移から作った方程式を解く
  各状態から始まる受理される言語を方程式で表す

正則表現の方程式：基本定理１（１）

- A, B, X : Σ 上の正則表現
  ただし，$\epsilon \notin L(A)$ とする
  $X = AX + B \;\Leftrightarrow\; X = A^*B$

- $X = A^*B \;\Rightarrow\; X = AX + B$
  $X = A^*B$ を代入
  $A^*B = AA^*B + B = A^+B + B = A^*B$

正則表現の方程式：基本定理１（２）

- $X = AX + B \;\Rightarrow\; X = A^*B$
  上より，$L(A^*B) \subseteq L(X)$ は明らか
  $L(X) - L(A^*B) \neq \emptyset$ と仮定（背理法）
  w : $L(X) - L(A^*B)$ の最短の語
  $w \notin L(A^*B)$ より，$w \notin L(B)$
  $X = AX + B$ より，$w \in L(AX)$
  $w = \alpha x \quad (\alpha \in L(A),\ x \in L(X))$
  $\alpha \neq \epsilon$ より，$|x| < |w|$
  w が $L(X) - L(A^*B)$ の最短の語なので $x \in L(A^*B)$
  $w = \alpha x \in L(A) \cdot L(A^*B) \subseteq L(A^*B)$ 　矛盾

正則表現の方程式：基本定理１（注）

- A, B, X : Σ 上の正則表現
  $X = AX + B$
  $\epsilon \in L(A)$ ならば，解は一通りとは限らない

- $X = (a + \epsilon)X + b$
  $X = a^*b,\quad X = (a + b)^*$ はいずれも解

正則表現の方程式：基本定理２ (1)

- A, X : Σ 上の正則表現
  ただし，$\epsilon \notin L(A)$ とする
  $X = AX \;\Leftrightarrow\; X = \emptyset$

  - $X = \emptyset \;\Rightarrow\; X = AX$
    $X = \emptyset$ を代入
    $\emptyset = A\emptyset = \emptyset$

正則表現の方程式：基本定理２ (2)

- $X = AX \quad \Rightarrow \quad X = \emptyset$

  $X \neq \emptyset$ を仮定（背理法）

  w ： $L(X)$ に属する最短の語

  $w \in L(X) = L(AX)$ より，

  $w = \alpha x \qquad (\alpha \in L(A),\ x \in L(X))$

  $\alpha \neq \epsilon$ より，$|x| < |w|$

  $x \in L(X)$ より，

  w が $L(X)$ に属する最短の語に矛盾

正則表現の方程式：例１ (1)

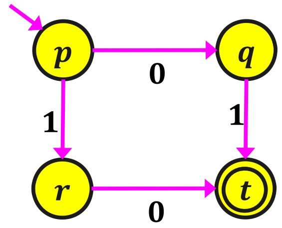
	0	1
$\to p$	q	r
q	−	t
r	t	−
$*\ t$	−	−

$p = 0q + 1r$

$q = 1t$

$r = 0t$

$t = \epsilon$ （受理状態）



p ：状態 p から受理状態に到達する語の集合を表す

$p = \{\alpha \in \Sigma^* \mid \delta(p, \alpha) \cap F \neq \emptyset\}$

$q,\ r,\ t$ も同様

正則表現の方程式：例１ (2)

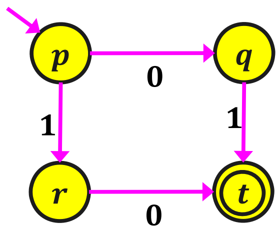
	0	1
$\to p$	q	r
q	−	t
r	t	−
$*\ t$	−	−

$p = 0q + 1r$

$q = 1t$

$r = 0t$

$t = \epsilon$ （受理状態）



$t = \epsilon$ を $q,\ r$ の方程式に代入： $q = 1,\ r = 0$

$q = 1,\ r = 0$ を p の方程式に代入： $p = 01 + 10$

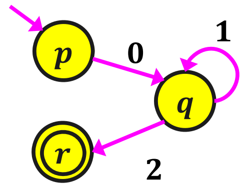
正則表現の方程式：例２

	0	1	2
$\to p$	q	−	−
q	−	q	r
$*\ r$	−	−	−

$p = 0q$

$q = 1q + 2r$

$r = \epsilon$ （受理状態）



$r = \epsilon$ を q の方程式に代入： $q = 1q + 2$

基本定理１： $q = 1^*2$

$q = 1^*2$ を p の方程式に代入： $p = 01^*2$

正則表現の方程式：例３

	0	1
$\rightarrow * p$	s	q
q	q	r
r	s	p
s	s	s

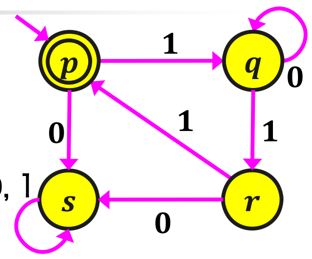
受理状態

$p = 0s + 1q + \epsilon$

$q = 0q + 1r$

$r = 0s + 1p$

$s = 0s + 1s = (0+1)s$



基本定理２： $s = \emptyset$

$s = \emptyset$ を r の方程式に代入： $r = 1p$

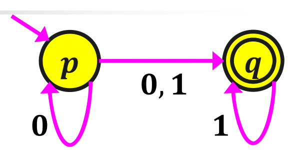
$r = 1p$ を q の方程式に代入： $q = 0q + 11p$

基本定理１： $q = 0^*11p$

$s = \emptyset$，$q = 0^*11p$ を p の方程式に代入： $p = 10^*11p + \epsilon$

基本定理１： $p = (10^*11)^*$

正則表現の方程式：例４



	0	1
$\rightarrow p$	p, q	q
$* q$	−	q

$p = 0p + 0q + 1q$

$q = 1q + \epsilon$ （受理状態）

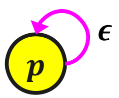
基本定理１： $q = 1^*$

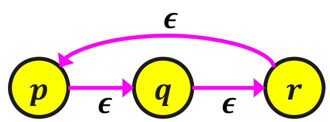
$q = 1^*$ を p の方程式に代入： $p = 0p + (0+1)1^*$

基本定理１： $p = 0^*(0+1)1^*$

基本定理を思い出してみると

- **基本定理１**
  A，B，X：Σ 上の正則表現．ただし，$\epsilon \notin L(A)$ とする．
  $X = AX + B \iff X = A^*B$
- **基本定理２**
  A，X：Σ 上の正則表現．ただし，$\epsilon \notin L(A)$ とする．
  $X = AX \iff X = \emptyset$
- **条件** $\epsilon \notin L(A)$
  - ϵ 動作だけで元の状態に戻らなければ満たす





本日の講義のまとめ

1. 有限オートマトン
2. 正則表現と正則言語
   - 正則表現
   - 有限オートマトンから正則表現へ
     - 動的計画法による方法
     - 正則表現の方程式による方法
3. 正則言語の性質
4. 文脈自由文法と言語
5. プッシュダウン・オートマトン
6. 文脈自由言語の性質
7. チューリングマシン

テキスト
3.1～3.2.2節