計算の理論 I
－ 正則表現 －

月曜３校時
大月 美佳

雑談

- 掲示板
  - http://www.cs.is.saga-u.ac.jp/lecture/automaton/sylpheed/
- その他
  - ミニテスト：もうちょっと待って

今日の講義内容

1. 前回の復習
   1. ε-動作を含むNFAからNFAへの変換
2. 今日の新しいこと
   1. 正則表現
      1. 正則表現を使うと嬉しいこと
      2. 前準備
      3. 正則表現の定義
      4. 正則表現の例

ε-動作なしNFAと ε-動作ありNFAの等価性 (p.34 定理2.2)

ε-動作ありNFA: M (Q, Σ, δ, q_0, F)が
ε-動作なしNFA: M´(Q, Σ, δ´, q_0, F´)
で模倣できる (定理2.2:証明→帰納法)。

F?? ?F ? {q_0}　　? - CLOSUREがFの元を含むとき
?F　　そうでないとき

? ?(q,a) ? ?̂(q,a)　(q ? Q, a ? ?)

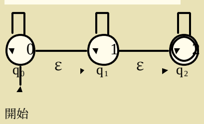
ε-動作ありNFAを模倣する ε-動作なしNFAの例 (p. 36 例2.9)

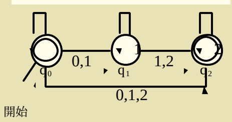
δ

	0	1	2	ε
q_0	$\{q_0\}$	○	○	$\{q_1\}$
q_1	○	$\{q_1\}$	○	$\{q_2\}$
q_2	○	○	$\{q_2\}$	○

δ´

	0	1	2
q_0	$\{q_0, q_1, q_2\}$	$\{q_1, q_2\}$	$\{q_2\}$
q_1	○	$\{q_1, q_2\}$	$\{q_2\}$
q_2	○	○	$\{q_2\}$



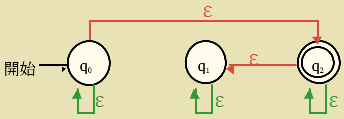


ε-動作を含むNFA→NFA 例1

- ε-動作を含むNFA
  ({q_0, q_1, q_2}, {0, 1}, δ, q_0, {q_2})
  δは下表

	0	1	ε
q_0	q_0, q_1	-	q_2
q_1	q_2	q_0, q_1	-
q_2	q_0, q_1	-	q_1

例1 ε-CLOSURE



	0	1	ε
q_0	q_0, q_1	-	q_2
q_1	q_2	q_0, q_1	-
q_2	q_0, q_1	-	q_1

ε-CLOSURE	ε
q_0	q_0, q_1, q_2
q_1	q_1
q_2	q_1, q_2

ε-動作を含むNFA→NFA 例2

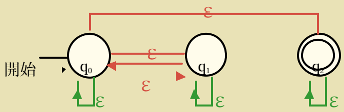
- ε-動作を含むNFA

$(\{q_0, q_1, q_2\}, \{0, 1\}, \delta, q_0, \{q_2\})$

δは下表

	0	1	ε
q_0	q_0	-	q_1
q_1	-	q_0, q_2	q_0
q_2	q_1	q_2	q_0

例2 ε-CLOSURE



	0	1	ε
q_0	q_0	-	q_1
q_1		q_0, q_2	q_0
q_2	q_1	q_2	q_0

ε-CLOSURE	ε
q_0	q_0, q_1
q_1	q_0, q_1
q_2	q_0, q_1, q_2

今日の新しいこと

- 正則表現
  1. 正則表現を使うと嬉しいこと
  2. 前準備
  3. 正則表現の定義
  4. 正則表現の例

正則表現を使えると嬉しいこと

- パターンマッチにしょっちゅう使う
  - UNIXのシェル
    - sh, csh, ksh, tcsh, bash
    - ファイルの名前
  - UNIXのコマンド、プログラミング言語
    - egrep, awk, sed, perl
    - ファイルの中の文字列の処理

Bash のファイル名展開

% ls

aaa, aab, aac, aba, aca, ada, abc, bbb, bcb

% echo a[abc]a

aaa aba aca

Perl での処理

Sun May 27 21:51:40 JST 2001

↓

s/¥w+ (¥w+) (¥d¥d) ¥d¥d:¥d¥d:¥d¥d ¥w+ (¥d+)/Today is ¥1 ¥2, ¥3./;

↓

Today is May 27, 2001.

前準備 その1
(記号列の集合)

- アルファベット Σ
- Σ 上の記号列 Σ＊
- Σ＊の部分集合 L, L_1, L_2

a, b, c

Σ

Σ＊

ε , a, b, c, ab, ac, ba, bc, ca, cb, aab, aba, ….

L

ε , ab, bc, aba, bcca

L_1

a, bc, aca

ba, abc

L_2

前準備 その2
(記号列の集合の演算)

- $L_1 L_2$ ={xy | x∈L_1, y∈L_2 }←連接

L_1

a, bc, aca

L_2

ba, abc

$L_1 L_2$

aba, bcba, acaba, aabc, bcabc, acaabc

前準備 その3
(記号列の集合の演算)

L^0 ? {?}

L^i ? $LL^{i?1}$ (i ? 1)

L^0

ε

L^1 ? L

ba, abc

L^2 ? LL^1 ? LL

baba, baabc, abcba, abcabc

L^i ? $LL^{i?1}$? LL...L

ba…ba, ba…abc, … …, abc…abc

前準備 その4
(Kleene閉包)

L^* ? $\bigcup_{i?0}^{?} L^i$ Lの*Kleene*閉包 (closure)（または単に閉包 (closure)）

? L^0 ? L^1 ? ...? L^i ? ...? $L^?$

L^0

ε ∪

L^1

ba, abc ∪…∪

$L^?$

ba…ba, ba…abc, … …, abc…abc

L^*

ε , ba, abc, …, ba…ba, ba…abc, …, abc…abc

前準備 その5
(正閉包)

$L^?$? $\bigcup_{i?1}^{?} L^i$ Lの正閉包 (positive closure)

L^1

ba, abc ∪…∪

$L^?$

ba…ba, ba…abc, … …, abc…abc

$L^?$が?を含むのはLが?を含むときのみ

$L^?$

ba, abc, …, ba…ba, ba…abc, …, abc…abc

教科書の例
(p. 37 例2.10)

L_1 ? {10, 1}

L_2 ? {011, 11}

$L_1 L_2$? {10011, 1011, 111}

$\{10, 11\}^*$? {?, 10, 11, 1010, 1011, 1110, 1111, ... }

$\{10, 11\}^?$? {10, 11, 1010, 1011, 1110, 1111, ... }

正則表現の定義
(p. 37)

1. ∅は正則表現で、その表す集合は空集合である。
2. εは正則表現で、その表す集合は{ε}である。
3. Σの各元aに対してaは正則表現で、その表す集合は{a}である。
4. rとsがそれぞれ言語RとSを表す正則表現のとき、(r+s)、(rs)、および(r*)は正則表現で、その表す集合はそれぞれ、R∪S、RS、R*である。

正則表現の例

- ∅、ε、a
- (∅+ε)=∅∪{ε}={ε}
- (∅+a)=∅∪{a}={a}
- (ε+a)={ε}∪{a}={ε, a}
- (a+b)={a}∪{b}={a, b}
- ∅ε={}{ε}={}=∅
- ∅a={}{a}={}=∅
- εa={ε}{a}={a}
- ab={a}{b}={ab}
- ∅*={}*=∅
- ε*={ε}*={ε}
- a*={a}*={ε, a, aa, aaa, ...}
- ((∅+ε)+a)=(ε+a)={ε}∪{a}={ε, a}
- (∅(a+b))={}{a, b}={}=∅
- ((ab)+ε)={ab}∪{ε}={ε, ab}
- ((ab)*)={ab}*={ε, ab, abab, ababab, ... }

正規表現の演算の強さ

* > 連接 > +

- ((0(1*))+0) → 01*+0
- (1+(10))* → (1+10)*
- ((1(1(1*)))+(01)) → (111*+01)

教科書の例 その1
(p. 38, 例2.11)

- 00={00}
- (0+1)*={0, 1}*
  ={ε, 0, 1, 00, 01, 10, 11, ...}
- (1+10)*={1, 10}*
  ={ε, 1, 10, 11, 110, 101, 1010, ... }
- (0+ε)(1+10)*={0, ε}{1, 10}*
  ={0, ε} {ε, 1, 10, 11, 110, 101, 1010, ... }
  = {ε, 0, 1, 01, 10, 010, 11, 011, 110, 0110, 101, 0101, 1010, 01010, ... }

教科書の例 その2
(p. 38, 例2.11)

- (0+1)*011={0, 1}*011
  ={ε, 0, 1, 00, 01, 10, 11, ...}{011}
  ={011, 0011, 1011, 00011, 01011, 10011, 11011, ... }
- 0*1*2*={0}*{1}*{2}* 図2.8のNFA
  ={ε, 0, 00, ... }{ε, 1, 11, ... } {ε, 2, 22, ... }
  = {ε, 0, 1, 01, 012, 00, 001, 0011, 0012, 00112, 001122, 000, 0001, 00011, 00012, 000111, 000112, 0001112, 00011122, 000111222, ... }

(1+10)*の性質

- 1で始まり、連続した0を含まない列か空列から成る集合

→帰納法で示す。

$(1 ? 10)^{*} ? \{1,10\}^{*} ? \square_{i ? 0}^{?} \{1,10\}^{i}$

$i ? n$のとき、$\square_{i ? 0}^{n} \{1,10\}^{i}$が

1で始まり連続した0を含まない列か空列のみから成るためには

$\{1,10\}^{0},\ldots,\{1,10\}^{n}$がそれぞれ

連続した0を含まない列か空列のみからなっていればよい。

帰納法での証明

1. $i ? 0$のとき、$\{1,10\}^{0} ? \{\varepsilon\}$
2. $i ? 1$のとき、$\{1,10\}^{1} ? \{1,10\}$

   で、1で始まり連続した0を含まない列のみを含む。
3. $i ?$ nのとき、

$\{1,10\}^{n}$が$\{1x_1,\ldots,1x_k\}$のように

1で始まり連続した0を含まない列のみを含む

とすれば($x_1,\ldots, x_k$は連続した0を含まない列)、

$i ? n ? 1$のとき、

$\{1,10\}^{n ? 1} ? \{1,10\}\{1,10\}^{n} ? \{11x_1,101x_1\ldots,11x_k,101x_k\}$

となり、1で始まり連続した0を含まない列のみを含むようになる。

よって1,2,3より、

$(1 ? 10)^{*}$は1で始まり連続した0を含まない列か空列のみを含む。

おまけ

- 正閉包に対応する正則表現

$rr^{*} ? r^{?} ? \square_{i ? 1}^{?} r^{i}$

$$\begin{aligned}
0^{*}1^{*}2^{*} &? 0^{?}1^{?}2^{?} \\
&? \{0, 00, \ldots\}\{1, 11, \ldots\}\{2, 22, \ldots\} \\
&? \{012, 0122, 0112, 01122, 0012, \\
&\quad 00122, 00112, 001122, \ldots\} \\
&? 0^{*}1^{*}2^{*}
\end{aligned}$$

今日のミニテスト

- ミニテスト
  - 教科書・資料を見ても良い
- 資料、ミニテストがない人は前へ
- 提出したら帰って良し
- 次回
  - 有限オートマトンと正則表現の透過性