創想館　14-201

2008-8

Tuesday　SPRING

13:00－14:30

2008年度

「情報工学」

第8回講義

前回学んだこと

復 習

- 覚えていなければいけないこと
- 答えられなければならないこと

✓「0」と「1」の個数が異なる‘01符号’によって表される‘文字’(A, B, C･･･など)で作られる文章は，各文字にどのような(長さの)符号が割り振られるかによって単位時間内に送られる情報に相違が生じる．

✓ 単位時間内に送れる情報量の最大値 $R_{max} = C$ (符号容量)は，十分に分に長い時間 T 秒内に送れる文字の組合せ総数を $N(T)$ とした場合，

$$C = \lim_{T\to\infty} \frac{\log_2 N(T)}{T} \quad ①$$

ここで $N(T)$ は，差分方程式 $N(T) = \sum_{i=1}^{N} N(T - \tau_i)$ ② を満たす．

ANIMATION

前回学んだこと

復 習

✓ 単位時間内に送れる情報量の最大値 $R_{max} = C$ (符号容量)は，十分に分に長い時間 T 秒内に送れる文字の組合せ総数を $N(T)$ とした場合，

$$C = \lim_{T\to\infty} \frac{\log_2 N(T)}{T} \quad ①$$

ここで $N(T)$ は，差分方程式 $N(T) = \sum_{i=1}^{N} N(T - \tau_i)$ ② を満たす．

この差分方程式を解くには，方程式 $\sum_{i=1}^{N} W^{-\tau_i} = 1$ ③ の最大根を求め，

$N(T) = AW^T$ として，

$$C = \lim_{T\to\infty} \frac{\log_2 AW^T}{T} = \lim_{T\to\infty}\left(\frac{\log_2 A}{T} + \frac{T \log_2 W}{T} \right) = \log_2 W \quad ④$$

を計算すれば，符号容量 C を得ることができる．

ANIMATION

復 習

前回学んだこと

✓ また，単位時間内に送れる情報量が最大となる時の‘各文字’（A，B，C・・・など）の出現頻度（生起確率）$p(\tau_i)$を求めるには，⑤式で求めたCの値を用いて，

$$p(\tau_i)=\frac{1}{A}2^{-\tau_i C} \quad ⑤$$

を計算すればよい．

ここでτ_iは，各文字（A，B，C・・・など）の長さ（各文字を表す‘01符号’「０」と「１」の個数によって決まる）．

猶，Aの値は，条件式 $\sum_{i=1}^{N} p(\tau_i)=1$ から決まる．

ANIMATION

復 習

Example：4.3.1

$$\begin{pmatrix} X \\ L \end{pmatrix}=\begin{pmatrix} A & B & C \\ 1 & 2 & 2 \end{pmatrix}$$ の時、C および p_i を求めてみよう．

意味するところ・・・

memo

各文字を表す 0,1 符号 ⇨ A：0，B：01，C：10

文章 ⇨ 01 0 1010 0 0110 0・・・・・

B A C C A B C A

☞ 実際には，文字間スペース，単語間スペースを考える必要がある．

ANIMATION

復 習

Example：4.3.1

$$\begin{pmatrix} X \\ L \end{pmatrix} = \begin{pmatrix} A & B & C \\ 1 & 2 & 2 \end{pmatrix}$$ の時、C および p_i を求めてみよう.

（解）　差分方程式は、

$$N(T) = N(T-1) + N(T-2) + N(T-2) \quad （1）$$

$$N(T) = AW^T \quad （2）$$

（2）→（1）　$AW^T = AW^{T-1} + AW^{T-2} + AW^{T-2}$

$\div AW^T$　$1 = W^{-1} + W^{-2} + W^{-2}$

$$\therefore \quad 2W^{-2} + W^{-1} - 1 = 0$$

ここで、$W^{-1} = X$ と置くと…

$$2X^2 + X - 1 = 0 \quad \Rightarrow \quad (2X-1)(X+1) = 0$$

よって、$X = \frac{1}{2} \ or \ -1$

復 習

$$\therefore \quad W = X^{-1} = 2 \ \ or \ \ -1$$

よって、差分方程式の一般解は、$N(T) = A_1(2)^T + A_2(-1)^T$ （3）

初期条件として、Table-3より、T=1、T=2の場合を用いる

$$N(1) = 1 、 N(2) = 3$$

これを、（3）式に代入すると、

$$1 = A_1 2^1 + A_2(-1)^1$$
$$3 = A_1 2^2 + A_2(-1)^2$$

整理すると…

$$2A_1 - A_2 = 1$$
$$4A_1 + A_2 = 3$$

$$\therefore \quad A_1 = \frac{2}{3} 、 A_2 = \frac{1}{3}$$

よって完全解は…　$N(T) = \frac{2}{3}2^T + \frac{1}{3}(-1)^T$ （4）

この式によって、任意の T に対する $N(T)$ を算出することが出来る！

ANIMATION

復 習

$\therefore W = X^{-1} = 2 \text{ or } -1$

T(sec)	文字の組合総数（文章）	N(T)
1	A	1
2	AA、B、C	3
3	AAA、BA、AB、CA、AC	5
4	AAAA、BAA、ABA、CAA、ACA、AAB、BB、CB、BAC、BC、CC	11
5	・・・・・・・・・・・・・	21
6		
7		
8		

Table.3

ANIMATION

復 習

ここで、$T \to \infty$ とすると、 $N(\infty) = \underbrace{\frac{2}{3}2^{\infty}}_{大} + \underbrace{\frac{1}{3}(-1)^{\infty}}_{1/3 \text{ or } -1/3} \Rightarrow \frac{2}{3}2^{\infty}$

よって無視できる！

よって、$W=2$ 、~~$W=-1$~~

これが､「最大根のみを採用すればよい！」という理由　！！

そこで、 $C = \lim_{T\to\infty} \frac{\log_2 N(T)}{T} = \lim_{T\to\infty} \frac{\log_2 AW^T}{T}$

$= \underbrace{\lim_{T\to\infty} \frac{\log_2 A}{T}}_{0} + \lim_{T\to\infty} \frac{\cancel{T}\log_2 W}{\cancel{T}}$

$= \log_2 W$

$= \log_2 2 \;= 1\, bit/sec$

復習

次に、この C を与える p_i を求めて見る.

$$p = \frac{1}{N(T)} = \frac{1}{AW^T} = \frac{1}{A}W^{-T} \Leftarrow C = log_2 W$$

$$= \frac{1}{A}(2^C)^{-T} = \frac{1}{A}2^{-CT} \Rightarrow \frac{1}{A}2^{-C\tau_i}$$

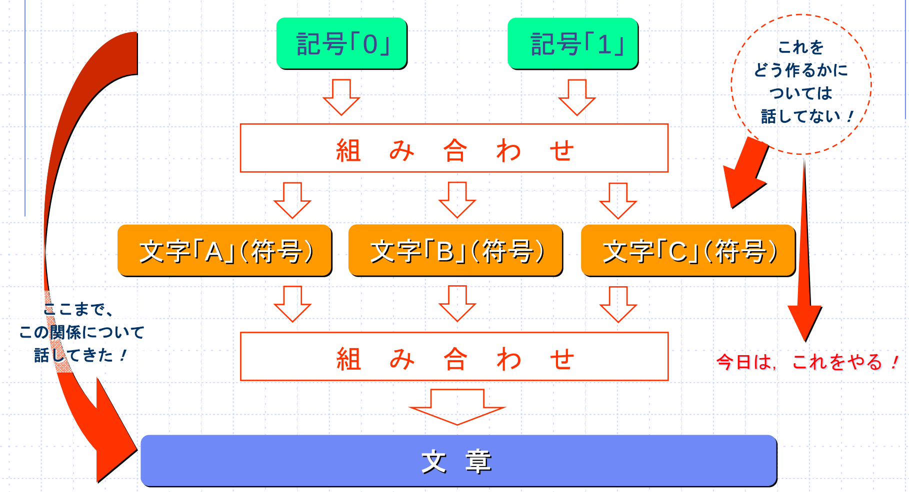
$$A: \quad p_i = \frac{1}{A}2^{-\tau_1 C} = \frac{1}{A}2^{-1\cdot1} = \frac{1}{2A} \Rightarrow 50\%$$

$$B: \quad p_i = \frac{1}{A}2^{-\tau_2 C} = \frac{1}{A}2^{-2\cdot1} = \frac{1}{4A} \Rightarrow 25\%$$

$$C: \quad p_i = \frac{1}{A}2^{-\tau_2 C} = \frac{1}{A}2^{-2\cdot1} = \frac{1}{4A} \Rightarrow 25\%$$

以上、A、B、C 各符号の生起確率が其々 50%、25%、25% の時、通信速度は最大となり、その時の符号容量は、1 (bit / sec) となることがわかった.

[確認] 「記号」・「符号」・「文章」の関係



§５　離散的情報の符号化

各文字の0,1符号を
どう決めるかの話

§５.１ 符号化の方針

（１）能率 → 冗長性低減

相反性

（２）確実性 → 冗長性増大

最適化 ⇒ 研究、学問

■ 「符号化法」とは　・・・

能率と確実性の最適化のための手法

■ 符号化の条件（整合条件）：

通報の生起確率と各符号の長さの兼ね合い
を与える条件

ANIMATION

符号化の定理

（１）単位時間当たりに送られる最大の通報数は、C / H（個）である.

（２）C / H（個）以下で送る符号化の方法は、常に存在する.

・・・歴史的意義大

§5.2 符号化法

§5.2.1 シャノンの符号化法

- 冗長度という観点からは、ベストな符号化法ではない
  したがって、現在は使われていない ・・・ただし、歴史的意義は大

シャノンの符号化法の手順

- 例により、具体的な例を用いて説明する
- 下表のような生起確率をもつ4つの文字に対する符号化を考える

文字	生　起　確　率
A	P_1=1/2
B	P_2=1/4
C	P_3=1/8
D	P_4=1/8

[STEP.1] 生起確率の順に並べる

$$
\begin{array}{ll}
A & p_1 = 1/2 \\
B & p_2 = 1/4 \\
C & p_3 = 1/8 \\
D & p_4 = 1/8
\end{array}
$$

[STEP.2] 生起確率の大きい方から、０ヶ、１ヶ、２ヶ・・・と加えて
P_1、P_2, P_3 ・・・を作る.

$$P_1 = 0$$
$$P_2 = p_1 = 1/2$$
$$P_3 = p_1 + p_2 = 1/2 + 1/4$$
$$P_4 = p_1 + p_2 + p_3 = 1/2 + 1/4 + 1/8$$

[STEP.3]　P_j を２進数で表す

$P_1 = 0$

$P_2 = p_1 = 1/2$

$P_3 = p_1 + p_2 = 1/2 + 1/4 = \underline{1/2^1 + 1/2^2}$

$P_4 = P_1 + p_2 + p_3 = 1/2 + 1/4 + 1/8 = \underline{1/2^1 + 1/2^2 + 1/2^3}$

$0 \Rightarrow 0.000$

$0.1 \Rightarrow 0.100$

$\underline{0.1 + 0.01} \Rightarrow 0.110$

$\underline{0.1 + 0.01 + 0.001} \Rightarrow 0.111$

[STEP.4]　下の不等式を満たす整数 m_j を決める

$$\log_2\left(\frac{1}{p_j}\right) \le m_j < \log_2\left(\frac{1}{p_j}\right) + 1$$

例えば、

・ $j = 1$ の時　$\log_2 \frac{1}{p_1} \le m_1 < \log_2 \frac{1}{p_1} + 1$　(p_1 ← 1/2)

$$\log_2 2 \le m_1 < \log_2 2 + 1$$

$$1 \le m_1 < 2 \qquad \therefore \quad m_1 = 1$$

・ $j = 2$ の時　$\log_2 \frac{1}{p_2} \le m_2 < \log_2 \frac{1}{p_2} + 1$　(p_2 ← 1/4)

$$\log_2 4 \le m_2 < \log_2 4 + 1$$

$$2 \le m_2 < 3 \qquad \therefore \quad m_2 = 2$$

・同様にして、　$m_3 = 3$、　$m_4 = 3$

[STEP.5]　前の２進数の少数点以下 m_j 桁までを、最終符号として採用する.

通報	p_j	P_j	m_j	(最終)符号
A	1 / 2	0.000	1	0
B	1 / 4	0.100	2	10
C	1 / 8	0.110	3	110
D	1 / 8	0.111	3	111

小数(分数)の2進数化について

まず、10進数の2進数化(変換)法、一般論から

① 元の数字(10進数)を、答えが 1 になるまで 2 で割っていく.

② 商(答え)が「1」になったら、その「1」と各計算での余りを下から並べていく.(各計算に余りがなければ「0」を入れておく)

・・・これで、2進数表現となる

例えば

```
2 ) 22 ・・・0
2 ) 11 ・・・1
2 )  5 ・・・1
2 )  2 ・・・0
     1
```

⇒ 10110

検証

$$1\times 2^4 + 0\times 2^3 + 1\times 2^2 + 1\times 2^1 + 0\times 2^0$$
$$= 16 + 0 + 4 + 2 + 0 = 22$$

小数(分数)の2進数化について

1/3という10進数(分数)の2進数表現(分数)は 1/11

```
      0.0101・・・・・・・
11 ) 1
     100
      11
       100
        11
        ・・・
```

⇒ 0.0101・・・・・

割切れない!

理由 10進数の 0.1 は、2進数では正確に表すことができない!

近似値:0.00011001100110011001100 ...

これを 10進数で表すと

0.0999990463256835937 5...

0.1 とはならない!

10進数でも、1/3 を小数で正確に表せないようなもの

小数(分数)の2進数化について

1／3という10進数(分数)の2進数表現(分数)は1／11

```
         0.0101・・・・・・・      ⇨   0.0101・・・・・
    11 ) 1
         100
          11
         ----
           100
            11
            ・・・
```

割切れない！

理由　10進数の 0.1 は、2進数では正確に表すことができない！

近似値：0.0001100110011001100110 ...

このような数を含む問題は、試験にださないことにする！！

0.1 とはならない！

10進数でも、1/3 を小数で正確に表せないようなもの

Example：5.2.1

先の問題において，無作為な符号化の結果とシャノンの符号化の結果を符号化の効率で比較せよ．

⇨ 要するに、０と１と送るのにかかる時間 τ_0 および τ_1 をともに１secとし、各文字 A、B、C、D の生起確率がそれぞれ、１／２、１／４、１／８、１／８の時につき、無作為な符号化による符号００、０１、１０、１１とシャノンの符号化による符号０、１０、１１０、１１１から得られる通信速度Rを求め、符号容量との比 R／C を計算する。

必ず、覚えとけよ！

言葉の定義

符号化の効率： $\eta = \dfrac{R}{C}$　　　符号化の冗長度： $1-\eta = 1-\dfrac{R}{C}$

ANIMATION

【略解】

（１）シャノンの符号化法を用いた場合

	p_i	符号	長さ
A	1/2	0	1
B	1/4	10	2
C	1/8	110	3
D	1/8	111	3

$$N(T) = N(T-1) + N(T-2) + 2N(T-3)$$

$$1 = W^{-1} + W^{-2} + 2W^{-3}$$

$$2X^3 + X^2 + X - 1 = 0 \xRightarrow{\text{最大根}} X = 1/2 \Rightarrow W = 2$$

$$C = \log_2 W = \log_2 2 = 1(bit/\sec)$$

$$R = \frac{-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - 2\times\frac{1}{8}\log_2\frac{1}{8}}{\frac{1}{2}\times 1 + \frac{1}{4}\times 2 + 2\times\frac{1}{8}\times 3} = 1(bit/\sec)$$

ANIMATION

$$\eta = \frac{R}{C} = \frac{1}{1} = 1 \Rightarrow$$ 符号化の効率 100%

（2）無作為に符号化法した場合

	p_i	符号	長さ
A	1/2	00	2
B	1/4	01	2
C	1/8	10	2
D	1/8	11	2

$$N(T) = 4N(T-2)$$

$$1 = 4W^{-2}$$

$$4X^2 - 1 = 0 \xRightarrow{\text{最大根}} X = 1/2 \Rightarrow W = 2$$

$$C = \log_2 W = \log_2 2 = 1(bit/\sec)$$ シャノンと同じ

ANIMATION

$$R = \frac{-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - 2\times\frac{1}{8}\log_2\frac{1}{8}}{\frac{1}{2}\times 2 + \frac{1}{4}\times 2 + 2\times\frac{1}{8}\times 2} = \frac{7}{8} = 0.87\ (bit\ /\sec)$$

$$\eta = \frac{R}{C} = \frac{\frac{7}{8}}{1} = 0.87$$

⇒ 符号化の効率 87%

ANIMATION

2008年度

「情報工学」

第8回講義　おわり