創想館　14-201

2008-7

Tuesday　SPRING

13:00ー14:30

2008年度

「情報工学」

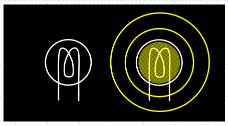
第7回講義

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

✓ 1個の電球(バイナリ信号発生器)からなるインジケータが送信できるエントロピーの最大値は，1(bit) となる．



$$\because \quad H_{\max} = \log_2 N = \log_2 2 = 1 \ (bit)$$

▶ したがって，電球が3個ならば，

$H_{\max} = \log_2 8 = \log_2 2^3 = 3 \ (bit)$　ということになる．

✓ 通信速度 R

$$R = \frac{H}{\tau} = \frac{-\sum_{i=1}^{N} p_i \log_2 p_i}{\sum_{i=1}^{N} p_i \tau_i} \ (bit/sec)$$

➤ この R の値によって，通信路の性能を評価することが出来る！

ANIMATION

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

✓ 通信速度 R

$$R = \frac{H}{\tau} = \frac{-\sum_{i=1}^{N} p_i \log_2 p_i}{\sum_{i=1}^{N} p_i \tau_i} \ (bit/sec)$$

➤ この R の値によって，通信路の性能を評価することが出来る！

▶ モールス符号を例として，通信速度 R を求めてみた．

☞ 「トン」の割合が多い方が，通信速度 R が速かった．

通信速度最速 R_{max} となる「トン」,「ツー」の生起確率は？

本日のメインテーマ

ANIMATION

§ 4.3 符号容量

- 通信速度最速 R_{max} を与える「トン」と「ツー」の割合を求める.
- しかし, いつまでもモールス符号では, ‘古臭い講義’ の感じがするので,「0」,「1」符号に話しを戻すことにしよう.
- 「0」,「1」を組合せて ‘文字’ (A, B, C・・・など)が作られ, それら ‘文字’ の組合わせによって文章が作られると考える.

各文字を表す 0,1 符号 ⇨ A : 00, B : 01, C : 10

文章 ⇨ 0100101000011000・・・・・

B A C C A B C A

☞ 実際には, 文字間スペース, 単語間スペースが必要となる.

ANIMATION

§ 4.3 符号容量

- 通信速度最速 R_{max} を与える「トン」と「ツー」の割合を求める.
- しかし, いつまでもモールス符号では, ‘古臭い講義’ の感じがするので,「0」,「1」符号に話しを戻すことにしよう.
- 「0」,「1」を組合せて ‘文字’ (A, B, C・・・など)が作られ, それら ‘文字’ の組合わせによって文章が作られると考える.
- 「0」と「1」の長さ(送るのにかかる時間) は同じでも, ‘文字’ を作るのに必要な「0」と「1」の個数が異なれば, 各文字(A, B, C・・・など)の長さ(送るのにかかる時間)は異なるということになる.
- 非常に長い文章, または多くの文章の中に現れる各文字 (A, B, C・・・など) の割合(出現頻度＝生起確率)は一定になるであろうという前提で, 以下話しを続けていく.

Spring 2008 Prof.M.Nakajima KEIO Univ. Yagami

ANIMATION

§ 4.3 符号容量

- 通信速度最速 R_{max} を与える「トン」と「ツー」の割合を求める．
- しかし，いつまでもモールス符号では，‘古臭い講義’の感じがするので，「0」，「1」符号に話しを戻すことにしよう．

以下，
各文字の長さが異なるとき， 長い文章の中に各文字がどのような割合で生起する場合に，単位時間に送れる情報量が最大となるか？
また， その最大値はいかほどになるかを求めてみよう．

長さ（送るのにかかる時間）は異なるということになる．

- 非常に長い文章，または多くの文章の中に現れる各文字（A，B，C・・・など）の割合（出現頻度＝生起確率）は一定になるであろうという前提で，以下話しを続けていく．

方 針

簡単のため 2文字しかないものとしよう． 各文字の生起確率が p_1，p_2 ならば，

$$\frac{\partial R(p_1, p_2)}{\partial p_1 \partial p_2} \Rightarrow 0$$ となる p_1，p_2 を求めることに帰着する．

C ：符号容量（bit/sec）

$p_1 + p_2 = 1$， すなわち $p_2 = 1 - p_1$ だから p_1 を上式左辺に代入して計算すれば， $\dfrac{dR(p_1)}{dp_1}$ の極大値として，$R_{max} = C$ が求まるはず！

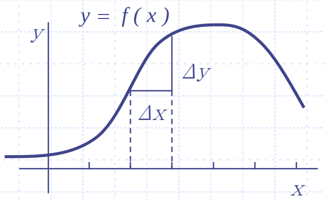
memo

微分方程式と差分方程式の関係

$$\frac{dy}{dx} = \frac{df(x)}{dx} \cong \frac{\Delta y}{\Delta x}$$

微分方程式の解

一般解： $y = a_1 e^{-\alpha x} + a_2 e^{-\beta x}$



以下、この種の問題に対する差分方程式の解き方（一般的な方法）を示す．

ANIMATION

[課 題]

T：十分に長い時間（秒）

$N(T)$ ：T 秒内に送れる文字の組合総数（文章）

τ_i：各文字(A, B, C・・・など)の長さ

一定の時間(T 秒)内に送れる情報量(文章の長さ)と
各文字の長さ(文字を表す符号における 0,1 の個数)
および各文字の出現頻度(生起確率)の関係を調べる

ANIMATION

[課 題]

T：十分に長い時間（秒）

$N(T)$ ：T 秒内に送れる文字の組合総数（文章）

τ_i：各文字(A, B, C・・・など)の長さ

ただし，簡単化のため，文字は３つ(A, B, C)のみとする．
（ 各文字の長さ：τ_a， τ_b， τ_c ）

この場合、T と $N(T)$ との関係は、線形差分方程式

$$N(T) = N(T-\tau_a) + N(T-\tau_b) + N(T-\tau_c) \qquad (1)$$

によって与えられる．

さ～て、では、この式を解いてみよう！

ANIMATION

[課 題]

線形差分方程式 $N(T) = N(T-\tau_a) + N(T-\tau_b) + N(T-\tau_c)$ を、より具体的に理解するため、３つの文字 Ａ、Ｂ、Ｃ の長さ τ_a、τ_b、τ_c を、それぞれ１秒、２秒、２秒として、その時の通信速度の最大値（符号容量）C 及び、その符号容量を与える文字 Ａ,Ｂ,Ｃの生起確率 p_a、p_b、p_c を求めてみる.

文字	A	B	C
長さ (sec)	1	2	2
p_i	p_a	p_b	p_c

Table.1

例えば，各文字A, B, Cに対応する「0」,「1」符号を
Ａ：0,　　Ｂ：01,　　Ｃ：10
とすれば，各文字の長さ（送るのにかかる時間）は，
Ａ：1秒,　　Ｂ：2秒,　　Ｃ：2秒
となる.

ANIMATION

[課 題]

線形差分方程式 $N(T) = N(T-\tau_a) + N(T-\tau_b) + N(T-\tau_c)$ を、より具体的に理解するため、３つの文字 Ａ、Ｂ、Ｃ の長さ τ_a、τ_b、τ_c を、それぞれ１秒、２秒、２秒として、その時の通信速度の最大値（符号容量）C 及び、その符号容量を与える文字 Ａ,Ｂ,Ｃの生起確率 p_a、p_b、p_c を求めてみる.

文字	A	B	C
長さ (sec)	1	2	2
p_i	p_a	p_b	p_c

Table.1

ここで、ある文章を各文字 Ａ, Ｂ, Ｃ の組み合わせによって送る場合を考えると、
１秒間以内に送れる組合せは、「Ａ」のみで、組合せの総数は１個しかないが、
２秒あれば、組合せは、「ＡＡ」,「Ｂ」,「Ｃ」で、組合せ総数は３個に増え、
３秒あれば、組合せは、「ＡＡＡ」,「ＡＢ」,「ＢＡ」,「ＡＣ」,「ＣＡ」となって、
組合せ総数は５個ということになる.

これを表にしてみると・・・

Table.2

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、AAC、BC、CC	11
5	・・・・・・・・・・・・・・・・	21

Table.2 が式（１）を満たしているかどうか、チェックしてみよう！

OK!

T= 3 時 $\underbrace{N(3)}_{5}=N(3-1)+N(3-2)+N(3-2)=\underbrace{N(2)}_{3}+\underbrace{N(1)}_{1}+\underbrace{N(1)}_{1}$

OK!

T= 4 時 $\underbrace{N(4)}_{11}=N(4-1)+N(4-2)+N(4-2)=\underbrace{N(3)}_{5}+\underbrace{N(2)}_{3}+\underbrace{N(2)}_{3}$

したがって、
T=5 時も、 $\underbrace{N(5)}_{\boxed{21}}=N(5-1)+N(5-2)+N(5-2)=\underbrace{N(4)}_{11}+\underbrace{N(3)}_{5}+\underbrace{N(3)}_{5}$

また、 $\underbrace{N(6)}_{\boxed{43}}=N(6-1)+N(6-2)+N(6-2)=\underbrace{N(5)}_{21}+\underbrace{N(4)}_{11}+\underbrace{N(4)}_{11}$

あとは同様に、 $N(7)=85$ $N(8)=171$ ・・・・・

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	・・・・・・・・・・・・・	43
7	・・・・・・・・・・・・・	85
8	・・・・・・・・・・・・・	171

Table.3

また、 $\underset{43}{N(6)} = N(6-1) + N(6-2) + N(6-2) = \underset{21}{N(5)} + \underset{11}{N(4)} + \underset{11}{N(4)}$

あとは同様に、 $N(7) = 85 \quad N(8) = 171$ ・・・・・

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	AAA… B…	11
5		…1
6		43
7	・・・	85
8	・・・・ ・・・・・	171

これで、差分方程式

$N(T) = N(T-\tau_a) + N(T-\tau_b) + N(T-\tau_c)$

が、本課題に反するものでないことを

確認することができた！

Table.3

ANIMATION

また、 $\underset{43}{N(6)} = N(6-1) + N(6-2) + N(6-2) = \underset{21}{N(5)} + \underset{11}{N(4)} + \underset{11}{N(4)}$

あとは同様に、 $N(7) = 85 \quad N(8) = 171$ ・・・・・

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
5		…1
6		43
7	・・・	85
8	・・・・ ・・・・・	171

では次に、差分方程式 $N(T) = N(T-\tau_a) + N(T-\tau_b) + N(T-\tau_c)$

を、一般的に解いてみよう・・・

$N(T) = N(T-\tau_a) + N(T-\tau_b) + N(T-\tau_c)$

が、本課題に反するものでないことを

確認することができた！

Table.3

では次に、差分方程式 $N(T)=N(T-\tau_a)+N(T-\tau_b)+N(T-\tau_c)$ を、一般的に解いてみよう・・・

微分方程式 ⇨ y'、y''、y''' 等からなる多項式.

したがって、一つの解の形として、$y=ae^{-\alpha x}$ を考え、

一般解（ $y=a_1e^{-\alpha x}+a_2e^{-\beta x}$ ）

特殊解、完全解という順で求めていく.

差分方程式の場合は、

１つの解の形として、$\boxed{N(T)=AW^T}$ を考える.

これを上式に代入すると・・・

$$Aw^T=AW^{T-\tau_1}+AW^{T-\tau_2}+AW^{T-\tau_3}$$

$R_{max}=C$

[両辺] $\div AW^T$

$$1=W^{-\tau_1}+W^{-\tau_2}+W^{-\tau_3} \quad (２)$$

ここで一時止めておいて、ここで出てくる W と符号容量 C の関係について考えてみる.

ANIMATION

W と C の関係

１つの文字が送られる確率 ： $p=\dfrac{1}{N(T)}$

その文字１つがもつ情報量 ： $log_2\dfrac{1}{p}=log_2 N(T)$

T ⇒ 大、 $N(T)$ ⇒ 大、 $log_2 1/p$ ⇒ 大

T	$N(T)$	$-log_2p$	$(-log_2p)/T$
1	1	0.000	0.000
2	3	1.585	0.792
3	5	2.322	0.774
4	11	3.459	0.865
5	21	4.392	0.878
6	43	5.426	0.904
7	85	6.409	0.916
8	171	7.418	0.927

Tが大きくなるにつれて、確かに $-\dfrac{log_2 p}{T}$ が大きくなる傾向あり！

$-\dfrac{log_2 p}{T}$ ⇨ 単位時間当たりの情報量 ⇨ したがって、その最大値が符号容量 C これを求める！

ANIMATION

よって、
$$C=\lim_{T\to\infty}\frac{-\log_2 p}{T}$$
$$=\lim_{T\to\infty}\frac{\log_2 N(T)}{T}$$

ここで、 $N(T)=AW^T$ を代入すると.

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

0

よって、 $\boxed{C=\log_2 W}$ (3)

よって、式（2）より W を求め、それを式（3）に代入することによって、符号容量 C を得ることが出来る.

では引き続いて、この C を与える各文字の生起確率 p を求めてみよう.

まず、C を与える生起確率 p を求める

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

また、 $C=\log_2 W$ より、 $W=2^C$

よって、 $p(T)=\dfrac{1}{A}\left(2^C\right)^{-T}=\dfrac{1}{A}2^{-CT}$

よって、 $T\to\tau_i$ とすると、

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

となり、これが符号容量 C（通信速度の最大値）を与える生起確率という
ことになる.

A は、あとで $\sum_{i=1}^{N}p(\tau_i)=1$ となるように調整！

[まとめ]

C の求め方

N 個の文字があり、各文字の長さが τ_1、τ_2、・・・ τ_N であるとした時、

$$W^{-\tau_1}+W^{-\tau_2}+\cdot\cdot\cdot+W^{-\tau_N}=1$$

なる方程式の根を求め、それを

$$C=\log_2 W$$

に代入することにより、Cを求める. ただし、根が２つ以上ある場合は、最大根を採用する.

また、C を与える各文字の生起確率 p については、

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

に τ_i と C を代入することによって求める.

ANIMATION

[まとめ]

ただし、これだけ丸暗記しても駄目！

C の求め方

N 個の文字があり、各文字の長さが τ_1、τ_2、・・・ τ_N であるとした時、

$$W^{-\tau_1}+W^{-\tau_2}+\cdot\cdot\cdot+W^{-\tau_N}=1$$

このあとの例題で、説明する！

なる方程式の根を求め、それを

$$C=\log_2 W$$

に代入することにより、Cを求める. ただし、根が２つ以上ある場合は、最大根を採用する.

また、C を与える各文字の生起確率 p については、

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

に τ_i と C を代入することによって求める.

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、$X = \dfrac{1}{2} \ or \ -1$

ANIMATION

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

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-2より、T=1、T=2の場合を用いる

$$N(1) = 1 \text{、} N(2) = 3$$

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…

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

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

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

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

大

1/3 or -1/3

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

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

0

$$=\log_2 W$$

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

ANIMATION

次に、この C を与える p_i を求めて見る.

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

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

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

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

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

以上、A、B、C 各文字の生起確率が其々 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.

ANIMATION

[宿題]

① 「Ａ」、「Ｂ」、「Ｃ」なる３つの文字がある.
各文字の長さが、それぞれ１秒、２秒、３秒である時、通信速度が最大となるのは、各文字の生起確率が如何程になる時か？
ただし、文字間のスペースは、どの文字の場合も同じで、文字の長さに含めて考えるものとする.

② 二つの温度計A、Bがある.
各温度計の応答時間および測定精度は、下表の通り、また両者の測定範囲は共に２０ ℃～５２ ℃である.
温度計A、Bは、どちらが多くの情報を送れるか.

温度計名	応答速度	測定精度
A	5 sec	3 ℃
B	60 sec	1 ℃

ANIMATION

2008年度

「情報工学」

第7回講義　おわり