東京大学大学院工学系研究科　電気系工学専攻　入試
平成 20 年　物理・情報　第 9 問

(1)

$$
\begin{aligned}
H(S) =& -p(S_0)\log_2 p(S_0) - p(S_1)\log_2 p(S_1) - p(S_2)\log_2 p(S_2) - p(S_3)\log_2 p(S_3) \\
&- p(S_4)\log_2 p(S_4) - p(S_5)\log_2 p(S_5) - p(S_6)\log_2 p(S_6) \\
=& \frac{5}{32}\log_2\frac{32}{5} + \frac{1}{16}\log_2 16 + \frac{1}{8}\log_2 8 + \frac{11}{32}\log_2\frac{32}{11} \\
&+ \frac{1}{8}\log_2 8 + \frac{5}{32}\log_2\frac{32}{5} + \frac{1}{32}\log_2 32 \\
=& 2.52
\end{aligned}
$$

エントロピー

定義は、

$$H = -\sum_i^n p(i)\log_2 p(i)$$

最大値は、全ての確率が等しいとき、つまり $p(i) = 1/n$ の時で、

$$H_m = \log_2 n$$

冗長度の定義、

$$(\text{冗長度}) = 1 - \frac{H}{H_m}$$

マルコフ過程では、単純にこのエントロピーの定義で求めることができないので注意。今回は「記憶のない情報源」という条件が付いているので問題ない。

マルコフ過程のエントロピーは、

$$H = -\sum_{ij} p(i)p(j|i)\log p(j|i)$$

である。

エントロピーの和は、

$$H = \sum_j p_j H_j$$

である。マルコフ過程のエントロピーは、各状態においてエントロピーを計算し、その和をとったものである。

(2)

平均符号長を最少とするために、ハフマンの符号化法を用いと以下のように符号化される。

$$S_0 : 11$$

$$S_1 : 1010$$

$$S_2 : 011$$

$$S_3 : 00$$

$$S_4 : 100$$

$$S_5 : 010$$

$$S_6 : 1011$$

平均符号長 τ は、

$$\tau = 2 \cdot \frac{5}{32} + 4 \cdot \frac{1}{16} + 3 \cdot \frac{1}{8} + 2 \cdot \frac{11}{32} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{5}{32} + 4 \cdot \frac{1}{32}) = \frac{83}{32}$$

ハフマンの符号化法

1. シンボルを確率の順に並べる。
2. 確率の低い 2 つのシンボルに 0 と 1 を当てる。
3. その 2 つのシンボルの確率を足し合わせて新しいシンボルとする。
4. 最初に戻って繰り返す。

平均符号長

平均通信時間と同様。
通報の長さ τ_j と、その生起確率 p_j より、

$$\tau = \sum_j p_j \tau_j$$

(3)

受信側が S_0 のシンボルを受信する確率は、

$$\begin{aligned} p^{'}(S_0) =& \frac{13}{16} \times p(S_0) + \frac{1}{32} \times p(S_1) + \frac{1}{32} \times p(S_2) + \frac{1}{32} \times p(S_3) \\ &+ \frac{1}{32} \times p(S_4) + \frac{1}{32} \times p(S_5) + \frac{1}{32} \times p(S_6) \\ =& \frac{13}{16} \times p(S_0) + \frac{1}{32} \times \{1 - p(S_0)\} \\ =& \frac{157}{1024} \end{aligned}$$

同様にして、

$$p^{'}(S_1) = \frac{41}{512}$$

$$p^{'}(S_2) = \frac{33}{256}$$

$$p^{'}(S_3) = \frac{307}{1024}$$

$$p^{'}(S_4) = \frac{33}{256}$$

$$p^{'}(S_5) = \frac{157}{1024}$$

$$p^{'}(S_6) = \frac{57}{1024}$$

(4)

シャノンの定理によれば、情報源のエントロピー H と、通信路容量 C に対して、$H-C$ まであいまい度を低くする符号が存在する。そのような符号を用いることができれば、あいまい度を限界まで下げることができる。

具体的には、符号にパリティビットを加える方法がある。これにより符号の冗長度は上がるが、誤りや損失を検出することができる。

シャノンの第 1 定理

情報源のエントロピーを H、通信路容量を C とする。
(1) この情報源からは単位時間当たり $\frac{C}{H}$ 個の通報数より多くは送れない。
(2) $\frac{C}{H}$ 個以下で送る符号化の方法は常に存在する。

シャノンの第 2 定理

エントロピー H の情報源を、通信路容量が C であり誤りのある系に接続したとき、
(1)$H \leq C$ であれば、誤りの確率をいくらでも 0 に近づける符号化法が存在する。
(2)$H > C$ であれば、そのような符号化はできない。
(3)$H > C$ の時は、あいまい度を $H-C$ にいくらでも近づけることができるが、それ以上は小さくできない。