

レポート問題 提出期限 2007年5月12日(木)

学籍番号 \_\_\_\_\_ 氏名 \_\_\_\_\_

ある病院では、風邪を引いているかどうかみてもらう患者がよくやってくる。病院では、まず熱があるかどうか調べ、次いでもっと精密な検査をすることになっている。熱を調べるだけで、どのくらいの情報量が得られるであろうか。  $X = \{\text{風邪である}, \text{風邪ではない}\}$ ,  $Y = \{\text{熱がある}, \text{熱が無い}\}$  とするとき、 $(X, Y)$  の同時確率分布は次のように与えられている。

	熱がある	熱が無い
風邪である	0.4	0.1
風邪ではない	0.2	0.3

以下の問いに答えよ。ただし、対数の底は2とし、 $\log_2 3 = 1.585$ ,  $\log_2 5 = 2.322$  とする。

1. 結合エントロピー  $H(X, Y)$  を求めよ。
2. 条件付確率  $P_{X|Y}(x|y)$  の表を書け。
3.  $Y$  で条件づけした  $X$  の条件付エントロピー  $H(X|Y)$  を求めよ。
4. 相互情報量  $I(X, Y)$  を求めよ。

レポート問題 解答例 (提出日 2007年5月10日(木))

1. 結合エントロピー  $H(X, Y)$  を求めよ.

$$\begin{aligned} H(X, Y) &= -0.4 \log \frac{2}{5} - 0.3 \log \frac{3}{10} - 0.2 \log \frac{1}{5} - 0.1 \log \frac{1}{10} \\ &= 0.4(\log 5 - 1) + 0.3(\log 5 + 1 - \log 3) + 0.2 \log 5 + 0.1(\log 5 + 1) \\ &= \log 5 - 0.3 \log 3 = \underline{1.8465} \end{aligned}$$

2. 条件付確率  $P_{X|Y}(x|y)$  の表を書け.

$P_{XY}(x, y)$	熱がある	熱が無い
風邪である	$\frac{2}{3}$	$\frac{1}{4}$
風邪ではない	$\frac{1}{3}$	$\frac{3}{4}$

3.  $Y$  で条件づけした  $X$  の条件付エントロピー  $H(X|Y)$  を求めよ.

$$\begin{aligned} H(X|Y) &= - \sum_{x,y} p_{XY}(x, y) \log p_{X|Y}(x|y) \\ &= -0.4 \log \frac{2}{3} - 0.2 \log \frac{1}{3} - 0.1 \log \frac{1}{4} - 0.3 \log \frac{3}{4} \\ &= 0.3 \log 3 + 0.4 = \underline{0.8755} \end{aligned}$$

4. 相互情報量  $I(X; Y)$  を求めよ.

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) = h(0.5) - H(X|Y) \\ &= 1 - 0.8755 = \underline{0.1245} \end{aligned}$$

ただし,  $h(p)$  は 2 値エントロピー.

(別)  $H(Y) = h(0.6) = 0.9710$  を先に求めて,

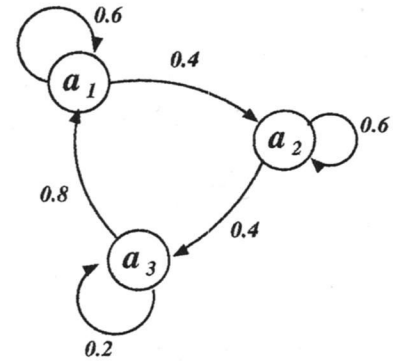
$$\begin{aligned} I(X; Y) &= H(X) + H(Y) - H(X, Y) \\ &= 1 + 0.9710 - 1.8465 = \underline{0.1245} \end{aligned}$$

レポート問題 提出期限 2007年6月7日(木)

学籍番号 \_\_\_\_\_ 氏名 \_\_\_\_\_

問1 あるマルコフ情報源  $S$  は、情報源アルファベットが  $A = \{a_1, a_2, a_3\}$  であり、右の図のような単純マルコフ連鎖をなしている。次の問いに答えよ。

- 1) 定常確率分布  $\mathbf{u} = (u_1, u_2, u_3)$  を求めよ。  
(ただし、 $u_i = \Pr(X_n = a_i)$  for  $i = 1, 2, 3$ )



- 2) このマルコフ情報源のエントロピー  $H(S)$  を求めよ。ただし、 $\log_2 3 = 1.585$ ,  $\log_2 5 = 2.322$  とする。

問2 文字列 “abracadabra” は、各アルファベット (a,b,c,d,r) の相対頻度が、それぞれ、 $\frac{5}{11}$ ,  $\frac{2}{11}$ ,  $\frac{1}{11}$ ,  $\frac{1}{11}$ ,  $\frac{2}{11}$  である。これを、出現確率  $P_X(x)$  ( $x \in \{a,b,c,d,r\}$ ) とみなす。このとき、

- 1) この情報源のエントロピーはいくらか。ただし、 $\log_2 11 = 3.459$  とする。

- 2) 2元ハフマン符号を構成せよ。

- 3) 変換後の文字列を示せ。また、平均符号長と1) で得られたエントロピーを比較せよ。