Title	2005年度 情報理論講義ノート
Author(s)	井上, 純一
Issue Date	2005-11-18T09:19:52Z
Doc URL	http://hdl.handle.net/2115/772
Rights(URL)	http://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Туре	learningobject
Note(URL)	http://www005.upp.so-net.ne.jp/j_inoue/index.html; http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/
Additional Information	There are other files related to this item in HUSCAP. Check the above URL.
File Information	InfoTheory05_slide5.pdf (第5回講義スライド)





情報理論 #5

第5回講義 5月23日

情報科学研究科 井上純一

http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/

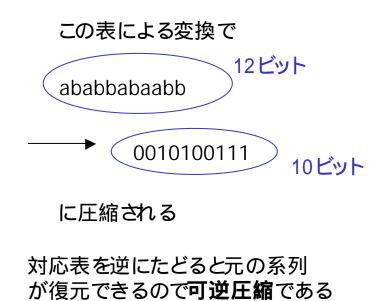


情報源圧縮

(先週の復習)

ababbabaabbb という系列をab|ab|ba|ba|aa|bb と2つずつに区切り、次の表に従って01の並びからなる符号に変換する

アルファベット	符号語	符号長		
aa	110	3		
ab	0	1		
ba	10	2		
bb	111	3		



復号

どの符号語も他の符号語の先頭部分になっていない **語頭条件**がみたされている



クラフト不等式

(先週の復習)

クラフト不等式

 $f: A \rightarrow B^+$ は一意復号可能であるとし、l(x)をxに対する符号語f(x)の長さとすると

$$\sum_{x \in A} K^{-l(x)} \le 1$$

が成り立つ。 KはBの要素数である。

(適用例)

$$A = \{aa, ab, ba, bb\}, B = \{0,1\}$$
 $\mathbf{y}_2(aa) = 00, \mathbf{y}_2(ab) = 10, \mathbf{y}_2(ba) = 11, \mathbf{y}_2(bb) = 110$ $l(aa) = l(ab) = l(ba) = 2, l(bb) = 3, K = 2$ のときは

$$\sum_{x \in A} K^{-l(x)} = 2^{-2} + 2^{-2} + 2^{-2} + 2^{-3} = \frac{7}{8} < 1 \qquad - \text{意復号可能}$$



クラフト不等式の証明#1

(先週後回しにした部分)

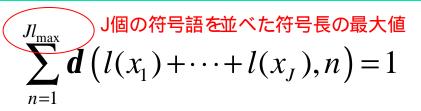
Jを任意の正の整数として、次が成り立つことに注意する

$$\left(\sum_{x \in A} K^{-l(x)}\right)^{J} = \sum_{x_1 \in A} \sum_{x_2 \in A} \cdots \sum_{x_J \in A} K^{-(l(x_1) + l(x_2) + \cdots + l(x_J))} \leftarrow$$

 $A_I(n)$ を全体の長さが n になるようなJ 個の符号語の並べ方の総数とすると

$$A_J(n) = \#\{x_1 x_2 \cdots x_J \in A^J : l(x_1) + l(x_2) + \cdots + l(x_J) = n\}$$

であり、ある符号語に対し、次が成り立つ



1をかける

クラフトの不等式の証明 #2

(先週後回しにした部分)

$$\left(\sum_{x \in A} K^{-l(x)}\right)^{J} = \sum_{n=1}^{Jl_{\max}} \sum_{x_1 \in A} \cdots \sum_{x_J \in A} d(l(x_1) + \cdots + l(x_J), n) K^{-n} = \sum_{n=1}^{Jl_{\max}} A_J(n) K^{-n}$$

 $\forall_{a,a} \in A$ に対して $f(a) \neq f(a')$ となるためには

$$A_J(n) \leq K^n$$
 Bの記号をn個並べてできる語の総数

であることが必要。

よって
$$\left(\sum_{x \in A} K^{-l(x)}\right)^{J} = \sum_{n=1}^{Jl_{\text{max}}} A_{J}(n) K^{-n} \le \sum_{n=1}^{Jl_{\text{max}}} 1 = Jl_{\text{max}}$$

$$\sum_{x \in A} K^{-l(x)} \le \left(J l_{\max} \right)^{1/J} = 1 \ (J \to \infty)$$



情報源符号の平均符号長

 l_i : 符号語 $f(a_i)$ の長さ

 p_i : 記号 a_i の出現確率 $p(a_i)$

$$L = \sum_{i=1}^{K} p_i l_i$$

平均符号長

例)

aa	00	1/4
ab	10	1/4
ba	01	1/4
bb	110	1/4

$$L = 2 \times (1/4) + 2 \times (1/4) + 2 \times (1/4) + 3 \times (1/4) = 9/4$$

平均符号長



平均符号長の下限

$$L - H_M(X) = -\sum_{i=1}^K p_i \log_M M^{-l_i} + \sum_{i=1}^K p_i \log_M p_i$$

平均符号長

情報源のエントロピー

$$= -\sum_{i=1}^{K} p_{i} \log_{M} cr_{i} + \sum_{i=1}^{K} p_{i} \log_{M} p_{i}$$

$$= \sum_{i=1}^{K} p_i \log \left\{ \frac{p_i}{r_i} \right\} - \log_M c \ge 0$$

確率の規格化条件

$$\sum_{i=1}^K r_i = c^{-1} \sum_{i=1}^K M^{-l_i} = 1$$
 より $c = \sum_{i=1}^K M^{-l_i} \le 1$ クラフト不等式より

$$L \ge H_{\scriptscriptstyle M}(X)$$

 $L \geq H_{_M}(X)$ | -意復号可能な場合には、平均符号長をエントロピー

クラフト不等式からの考察

記号*a*_iを符号化したものの長さを次のようにおく

ものの長さを次のようにおく
$$\sum_{i=1}^K M^{-l_i} \leq \sum_{i=1}^K M^{-\left|\log_M \frac{1}{p_i}\right|}$$

$$\leq \sum_{i=1}^K M^{-\log_M \frac{1}{p_i}} = 1$$

$$\int_{D_i}^{D_i} \log_M \frac{1}{p_i} \int_{D_i}^{D_i} \log_M \frac{1}{p_i} dx$$

括弧内以上の最小な整数

$$\log_{M} \frac{1}{p_{i}} \leq l_{i} < \log_{M} \frac{1}{p_{i}} + 1$$

$$\Rightarrow \sum_{i=1}^{K} p_{i} \log \frac{1}{p_{i}} \leq \sum_{i=1}^{K} p_{i} l_{i}$$

$$\Rightarrow \sum_{i=1}^{K} p_{i} \log \frac{1}{p_{i}}$$

$$\Rightarrow \sum_{i=1}^{K} p_{i} \log \frac{1}{p_{i}}$$