情報源 ${\cal S}$

情報の発生源

情報源Sのシンボル s_i

情報源Sが出す記号

情報源Sのアルファベット $S = \{s_1, \dots, s_q\}$

情報源Sの出すシンボルの集合

定常情報源

シンボルの出現確率は時間に依らず、常に一定な情報源

記憶のない情報源

あるシンボルの出現確率は、その前までに出現しているシンボルに依存しないような情報源

符号シンボル t_i

情報源Sを符号化する際に用いられるシンボル

符号アルファベット $T = \{t_1, \cdots, t_r\}$

符号シンボルの集合

基数r = |T|

符号シンボルの個数。rの符号アルファベットはr元符号と呼ばれる

語w Tのシンボルからなる有限列

語長, 長さ|w|

語を構成するシンボルの個数

空語 ϵ 語長が0の語

$$T^{+} = \bigcup_{i=1}^{n} \prod_{i=1}^{i} T$$

空語を含まない語の集合
$$T^* = \bigcup_{i=0}^n \prod_{j=1}^i T$$

空語を含んだ語の集合

符号 $\mathcal{C}:S \to T^*$

 $1, \dots, q$)}と書かれることもある

 $\mathcal{C}^*: S^* \to T^*$

Cを S^* から T^* の関数として拡張したもの。 同様に $C^*=\{w_{i_1}\cdots w_{i_n}\in T^*:w_{i_1},\cdots,w_{i_n}\in C\}$ と書か れることもある

符号語

符号の要素

平均符号長
$$L(\mathcal{C}) = \sum_{i=1}^q p_i |w_i|$$

符号Cにおける、語長の平均値

一意復号可能な符号

全単射な符号

ブロック符号

語長が全て同じ符号

語頭 ある語の先頭から初まっているシンボルの部分列のこと。例えば、語w=w'1のw'が語頭語頭符号

全ての符号語は語頭に他の符号語をもたないような符号

瞬時復号可能な符号

一意復号可能で、語を見たときに復元をすぐに行うことができる符号。語頭符号と同値

r元根付き木

空語 ϵ を根として、それに符号シンボル $t\in T$ を足した、 $\epsilon t=t$ という語を空語の子とし、さらに、その子にそれぞれ符号シンボル $t'\in T$ をたしたtt'という語を子にもつようにするという操作を、繰り返して得られる木。ただし、高さは有限

クラフトの不等式

符号長 l_1,\cdots,l_q を持つ瞬時符号は、高さが $\max\{l_1,\cdots,l_q\}$ のr元根付き木を考えたとき、符号語として選択された語の子、すなわち、その選択された節の語を語頭として持っているr元根付き木の語の数の割合が1を越えないときかつそのときのみ存在するという定理。すなわち、次の不等式を満たすときかつそのときのみ、語長 l_1,\cdots,l_q を持つ瞬時符号は存在する:

$$\sum_{i=1}^{q} \frac{1}{r^{l_i}} \le 1 \tag{1}$$

マクミランの不等式

瞬時符号と同じで、符号長 l_1,\cdots,l_q を持つ一意復号可能な符号の存在条件を伸べた定理。瞬時符号と同じで、条件式(1)を満たすときかつそのときのみ符号長 l_1,\cdots,l_q を持つ一意復号可能な符号は存在する。ここから、符号長が l_1,\cdots,l_q の一意復号可能な符号が存在するときかつそのときのみ同じ語長を持つ瞬時符号が存在することがすぐに従う

最適符号, コンパクト符号

基数と各情報源アルファベットの出現確率の分布が与えられたとき、平均符号長が最も小さくなる符号。任意の情報源は $r \geq 2$ とすれば、最適符号が必ず存在する

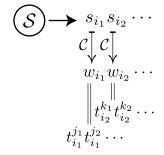


図1 情報源Sはシンボル列 $s_{i_1}s_{i_2}\cdots$ を出し、それを符号Cを利用して語の列 $w_{i_1}w_{i_2}\cdots$ へ符号化

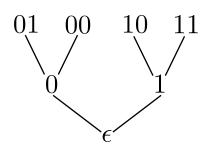


図2 高さが2の2元根付き木の例