

医用画像工学実験

－画像の情報量 編－

- 医療情報学
- 医用画像情報学
- 情報科学
-

「情報」という言葉をよく用いているが、そもそも情報とは何だろうか。

<目的>

- 画像の情報量を計算できるようになる.
- (医用)画像を情報量の側面から考察する.

情報ってなに？

- 英語 : information
- 以下, 何冊かの書籍からの抜粋
 - 同等な可能性を持ついくつかの事柄の中から特定の一つを指定するためのもの.
 - 対象物の状態が不確定である時, 状態を分類し他の可能な状態から区別することによって, 対象物の状態をより明確にするもの.
 - 生物の行動に影響を与えるすべての事象.

情報の量

- その情報を得た人にとっての内容の豊富さのこと

例:



13	14	15	16
9	10	11	12
5	6	7	8
1	2	3	4

入口 アパートのモデル図

Q. ケース1とケース2では、どちらが入口で得た情報の量が多い(豊富)と考えられるか？

A君は入口で次の情報を教えてもらった

情報: S君は11号室に住んでいる

- ケース1: A君はS君が3階に住んでいるのを知っていた.
- ケース2: A君はS君が何階に住んでいるのか知らなかった.

情報の量はどやって測る？

- 情報の量は、情報を得る前の可能性の数（前例では部屋の数）に関係し、その数が増すにつれ情報の量も増すようでないといけない（単調増加関数）。
- 情報の加法性が成り立たなければならない。
 - 情報1: S君の部屋は16室の11号室である。
 - 情報2: S君の部屋は3階にある。
 - 情報3: S君の部屋は左から3番目の部屋である。

– 「情報1」の量 = 「情報2」の量 + 「情報3」の量



- 情報の量は、可能性の数の対数で定義するのが便利
- 対数の底を2にとると、最小の情報（二択）の量は $\log_2 2 = 1$ となり、情報量の単位として都合がよい。

情報量の定義

p_i : ある出来事が発生する確率(事前確率)

自己情報量: I

$$I(p_i) = \log_2 \frac{\text{事後確率}}{\text{事前確率}}$$
$$= \log_2 \frac{1}{p_i} = -\log_2 p_i$$

平均情報量: H
(エントロピー)

$$H = p_1(-\log_2 p_1) + p_2(-\log_2 p_2) + \dots + p_N(-\log_2 p_N)$$

情報量の単位は [bit]

$$= -\sum_{i=1}^N p_i \log_2 p_i, \quad \text{ただし} \sum_{i=1}^N p_i = 1$$

最大エントロピー： H_{max}

すべての出来事が等確率で発生すると仮定した場合のエントロピー

最大エントロピーとエントロピーの違いは情報源に含まれる「無駄さ」である。

無駄さ＝冗長度 (Redundancy) : r は次のように定義される。

$$r = 1 - \frac{H}{H_{\max}} = 1 - \frac{\text{エントロピー}}{\text{最大エントロピー}}$$

データ圧縮

情報のムダな部分を取り除いて、データ量を減らすこと

↓
データの冗長性

↓
「情報量」と「データ量」の差

データ量 : データの物理的なサイズ

情報量 : データに含まれる情報の正味の価値

データ量と情報量

例： 100種類の文字からなる100文字の文字列

データ量

1文字 = 8ビット → $8 \times 100 = 800$ ビット (アルファベットなど)

1文字 = 16ビット → $16 \times 100 = 1600$ ビット (ひらがな, 漢字など)

情報量

データ量は変わっても, 情報の質は変わらない



「100文字の文字列である」

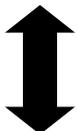
情報量は同じでも, データ化によってデータ量に差が生じる

情報量 < データ量

→ データの冗長性

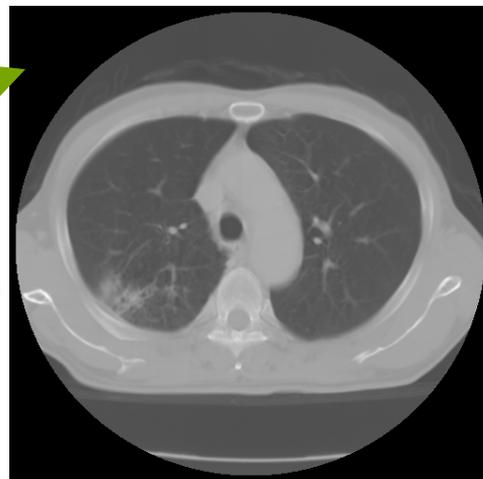
なぜ冗長性があるのか？

(答え) データの取り扱いが容易になる



(答え) 冗長性を持たないデータは，プログラムなどで取り扱う場合に処理効率が著しく悪化し，実用に適さない

冗長な部分
(背景, 寝台などの
身体情報以外のデータ)



情報量の数学的な定義

- ・情報量は数学的には確率で定義される。

確率が低い事象 → 情報量が多い

確率が高い事象 → 情報量が少ない

$$\text{平均情報量(エントロピー)} = -\sum p_i \log_2 p_i \quad (\text{bits})$$

例： 4つの値からなるデータ(A, B, C, D)

- ・出現率が均等である場合

$$p_A = p_B = p_C = p_D = 1/4$$

平均情報量 = 2.0 ビット

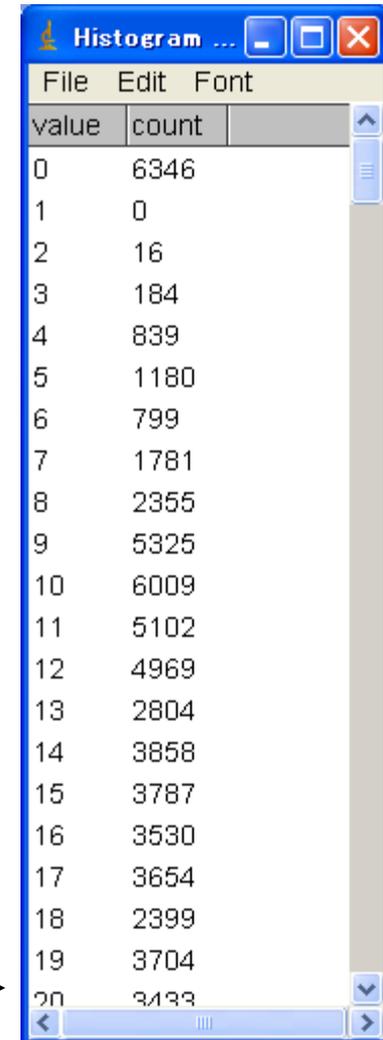
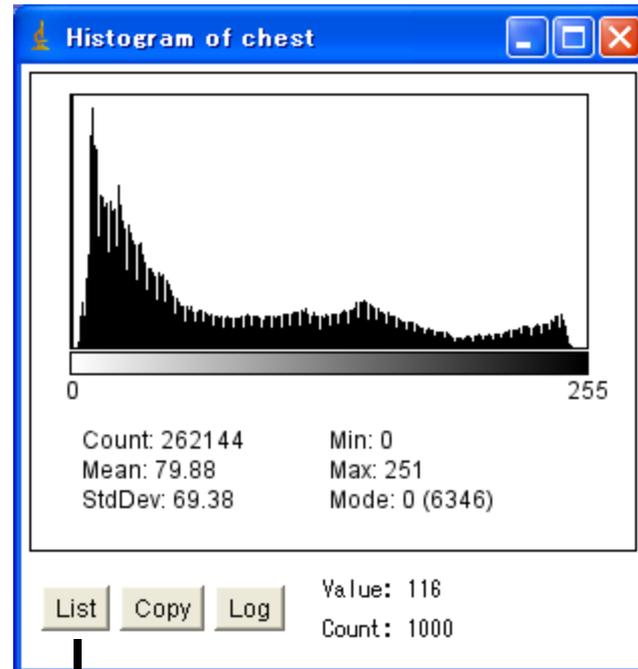
- ・出現率に偏りがある場合

$$p_A = 1/2, p_B = 1/4, p_C = p_D = 1/8$$

平均情報量 = 1.75 ビット

画像の情報量

- 画素値 i の発生頻度を p_i とする。
 - p_i はヒストグラムから得られる。



value	count
0	6346
1	0
2	16
3	184
4	839
5	1180
6	799
7	1781
8	2355
9	5325
10	6009
11	5102
12	4969
13	2804
14	3858
15	3787
16	3530
17	3654
18	2399
19	3704
20	3133