符号理論・暗号理論

- No.1 情報量 -

渡辺 裕

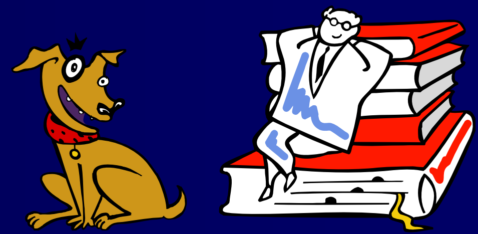
Coding Theory / Cryptography

- No.1 Information -

Hiroshi Watanabe

情報量

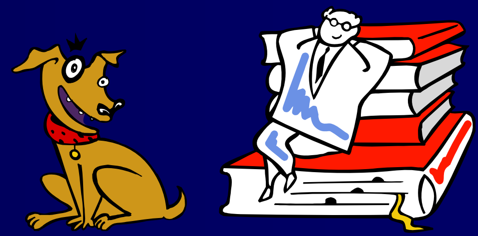
- 情報量とは？　…　情報の重要性を計りたい



  - 例1　　通報1A: “犬が教授にかみついた”
  　　　　通報1B: “教授が犬にかみついた”
  - 1Bは滅多に起こりそうにない → 情報としての価値が高い → 情報量が大きい
  - 例2　　通報2A: “その日は6月で雪だった”
  　　　　通報2B: “その日は6月で雨だった”
  - 2Aは滅多に起こりそうにない → 情報としての価値が高い → 情報量が大きい

Information Content

- What is it? … want to measure the importance of information



  - Ex.1 Message 1A: "A dog bite a professor."
    Message 1B: "A professor bite a dog."
  - 1B is not likely to happen → information value is high → information content must be large

  - Ex.2 Message 2A: "That day was snowing in June."
    Message 2B: "That day was raining in June."
  - 2A is not likely to happen → information value is high → information content must be large

情報量(2)

- 情報量として期待される性質１ ... 事象が起こる確率が低いほど大きな値をとる
  - 事象 x の出現確率 $p(x)$ が小さいほど、情報量 $I(x)$ が大きくなるような尺度が望まれる
  - $p(x)=1$ のときは、必ず起きる事象であり、情報としての価値がないから、$I(x)=0$ となるような尺度
  - $p(x)$ が0 に近いときは、ほぼ起こらない事象であり、情報としての価値が非常に高いから、$I(x)$ は∞に近づくような尺度

Information Content(2)

- Desired nature of information content (1) … Value should be large when event has low probability
  - Desired measure satisfies that information content $I(x)$ becomes large along with the occurrence probability $p(x)$ of an event x becomes small
  - Information content becomes $I(x)=0$ when $p(x)=1$ since it always occurs and there is no value as an information
  - Information content $I(x)$ approaches ∞ when $p(x)$ is close to 0 since such event seldom occurs

情報量(3)

- 情報量として期待される性質2 ... 独立した複数の事象の情報量は、個々の事象の情報量の和
  - 事象 x および事象 y が独立に起こったとき、それぞれの出現確率を $p(x), p(y)$、情報量を $I(x), I(y)$ とすると、事象 x と y の組み合わせに対する確率 $p(t)$ は $p(t)=p(x)p(y)$ であり、情報量 $I(t)$ は、$I(t)=I(x)+I(y)$
  - 例1と例2を組み合わせても、それぞれの事象... 例えば、“教授が犬に噛み付いた”、“その日は6月で雪だった” は独立

Information Content(3)

- Desired nature of information content (2) … Total information content of multiple independent events is given by summation of all event's information contents
  - Let the occurrence probabilities and information contents of the independent event x and y be $p(x), p(y)$ and $I(x), I(y)$. The combined probability $p(t)$ of the events x and y is $p(t)=p(x)p(y)$, and combined information content $I(t)$ should be an addition of two information contents $I(t)=I(x)+I(y)$
  - When Ex.1 and 2 are combined, each event is independent. i.e. "A professor bite a dog." is independent from "That day was snowing in June."

情報量(4)

- 情報量の定義
  - 情報量の性質1（単調減少性）および性質2（加法性）を満たす関数… 対数関数 (log)
  - 対数の底を2に選んだ情報量 $I(x)$ の定義… 出現確率 $p(x)$ である事象 x の自己情報量 (Self Information Content)

$$I(x) \equiv \log_2 \frac{1}{p(x)} = -\log_2 p(x)$$

  - 単位はbit: Binary Unit の略

Information Content(4)

- Definition of Information Content
  - Find function which satisfies desired nature (1) (Monotonic decreasing) and nature (2) (Addition) … Logarithmic function (log)
  - Definition of Information Content $I(x)$ having 2 for logarithmic base $I(x)$ … Self Information Content of event x with the occurrence probability $p(x)$

$$I(x) \equiv \log_2 \frac{1}{p(x)} = -\log_2 p(x)$$

  - Unit is bit: abbreviation of Binary Unit

復習

- 対数の演算
  - 対数の定義

$$\log_a x = y \quad \Leftrightarrow \quad x = a^y$$

  - 積の対数

$$\log_a xy = \log_a x + \log_a y$$

  - べき乗の対数

$$\log_a x^y = y \log_a x$$

Review

- Calculation of logarithm
  - Definition of logarithm

$$\log_a x = y \quad \Leftrightarrow \quad x = a^y$$

  - Logarithm of multiplied variables

$$\log_a xy = \log_a x + \log_a y$$

  - Logarithm of the variable to the power of other variable

$$\log_a x^y = y \log_a x$$

復習(2)

- 対数の計算
  - 情報理論でよく使う計算

$$-\log_a \frac{1}{x} = -\log_a x^{-1} = \log_a x$$

  - 底の変換

$$\log_a x = \frac{\log_b x}{\log_b a}$$

Review(2)

- Calculation of logarithm
  - Relation often used in information theory

$$-\log_a \frac{1}{x} = -\log_a x^{-1} = \log_a x$$

  - Change of base

$$\log_a x = \frac{\log_b x}{\log_b a}$$

情報量(5)

- 出現確率 $p(t)=p(x)p(y)$ を持つ事象 x と y の組み合わせ事象 t に対する自己情報量 $I(t)$ の加法性の確認

$$
\begin{aligned}
I(t) &= -\log_2 p(t) \\
&= -\log_2 p(x)p(y) \\
&= -\log_2 p(x) - \log_2 p(y) \\
&= I(x) + I(y)
\end{aligned}
$$

Information Content(5)

- Confirmation of additional characteristics of information content $I(t)$ where the combined event t (combined by x and y) has the occurrence probability $p(t)=p(x)p(y)$

$$
\begin{aligned}
I(t) &= -\log_2 p(t) \\
&= -\log_2 p(x)p(y) \\
&= -\log_2 p(x) - \log_2 p(y) \\
&= I(x) + I(y)
\end{aligned}
$$

計算例

- サイコロを1個振って1がでたときの情報量
- トランプのカードを1枚抜いたときにAであったときの情報量

Quiz

- How much is the information content when you get 1 by casting one dice?
- How much is the information content when you get ace by drawing one card?

平均情報量

- 自己情報量
  - 単一の通報のもつ情報量
- 平均情報量
  - ある出現確率で起こる通報の平均的な重要性の尺度
  - 重要性が異なる例
    - LA（6月）の天気予報
      - p(hazy sunshine)=0.95, p(shower)=0.05
      - ほとんど晴れなので、平均的には天気予報は重要ではない
    - 東京（6月）の天気予報
      - p(晴れ)=0.6, p(時々雨)=0.4
      - 晴れか雨かわからないので、天気予報は重要

Mean Information Content

- Self Information Content
  - Information Content of one message

- Mean Information Content
  - Measure of mean importance of messages occur in a certain probability
  - Example of different importance
    - Weather forecast of LA in June
      - p(hazy sunshine)=0.95, p(shower)=0.05
      - Weather forecast is not important in average since it is almost always hazy sunshine.
    - Weather forecast of Tokyo in June
      - p(fine)=0.6, p(occasionally rain)=0.4
      - Weather forecast is important since it is uncertain.

平均情報量(2)

- 平均情報量の定義
  - N 個の事象 $a_1, a_2, \dots ,a_N$ の出現確率を $p_1, p_2, \dots ,p_N$ としたときの平均情報量 I_{ave}

$$I_{ave} = -\sum_{i=1}^{N} p_i \log_2 p_i$$

  - 平均情報量はエントロピーとも呼ばれ、無記憶情報源 S に対するエントロピーを $H(s)$ と表記する

$$H(S) = I_{ave} = -\sum_{i=1}^{N} p_i \log_2 p_i$$

Mean Information Content(2)

- Definition of Mean Information Content
  - To N events $a_1, a_2, \ldots, a_N$ having occurrence probabilities $p_1, p_2, \ldots, p_N$, the mean information content I_{ave} is given as follows.

$$I_{ave} = -\sum_{i=1}^{N} p_i \log_2 p_i$$

  - Mean information content is called "Entropy." Entropy $H(S)$ for the memory-less source S is denoted by

$$H(S) = I_{ave} = -\sum_{i=1}^{N} p_i \log_2 p_i$$

平均情報量(3)

- 平均情報量の計算例
  - LA6月および東京6月の天気予報の平均情報量を $I_{ave}(L)$, $I_{ave}(T)$ とすると

$$I_{ave}(L) = -0.95\log_2 0.95 - 0.05\log_2 0.05 = 0.286$$

$$I_{ave}(T) = -0.6\log_2 0.6 - 0.4\log_2 0.4 = 0.997$$

となり、東京6月の天気予報の方が、平均情報量が大きく、重要な通報であることがわかる

Mean Information Content(3)

- Example of Mean Information Content
  - Let mean information content to weather forecast LA in July and Tokyo in July be $I_{ave}(L)$, $I_{ave}(T)$

$$I_{ave}(L) = -0.95\log_2 0.95 - 0.05\log_2 0.05 = 0.286$$

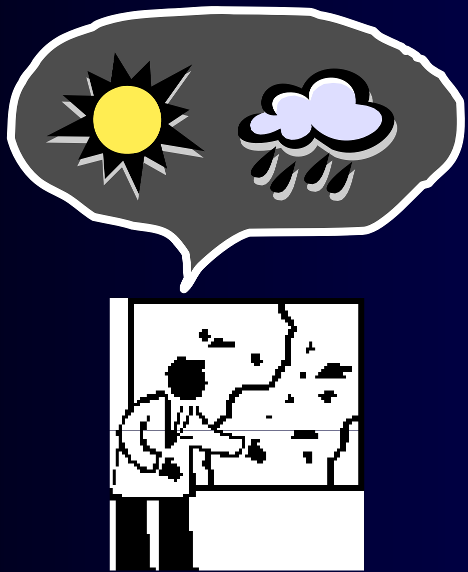
$$I_{ave}(T) = -0.6\log_2 0.6 - 0.4\log_2 0.4 = 0.997$$

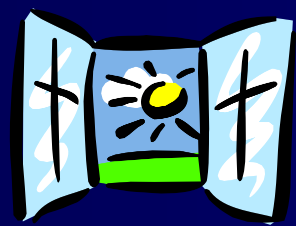
Thus, mean information content of weather forecast Tokyo in July has larger value, and is recognized more important message.

相互情報量

- 相互情報量の定義
  - 正確ではない情報を受けた場合の情報量の定義

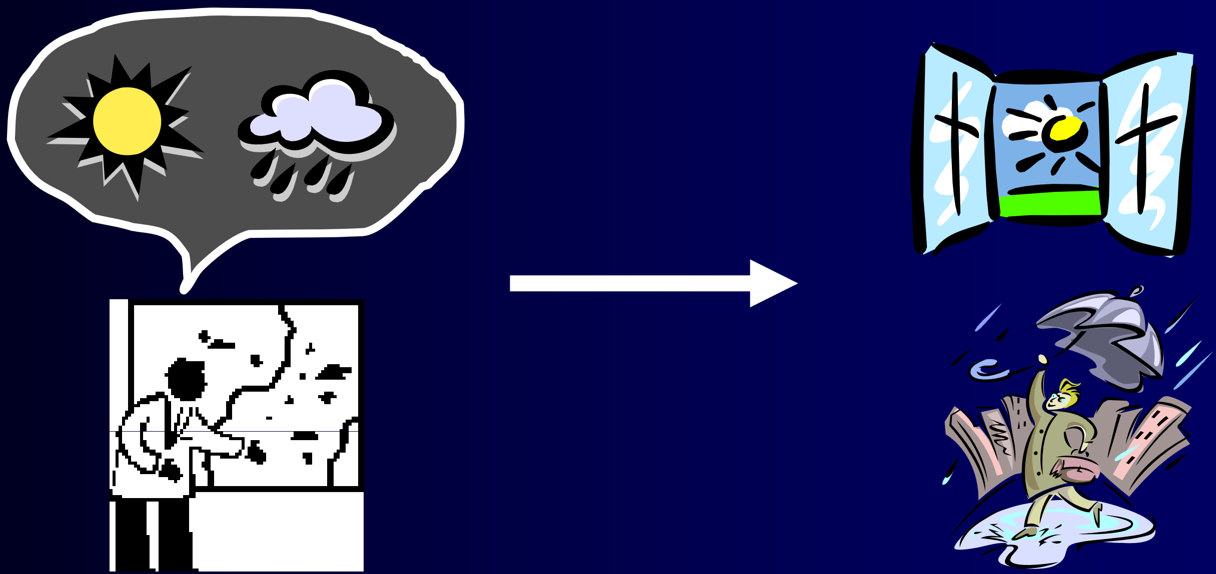
- 天気予報と実際の天気





Mutual Information

- Definition of Mutual Information
  - Define information content when receiving incorrect information
  - Weather forecast and real weather



相互情報量(2)

- 天気予報の例
  - X: 実際の天気
  - Y: 天気予報
  - $P(x,y)$: それぞれの確率
  - $P(x), P(y)$: 結合確率分布

$P(x,y)$		Y		$P(x)$
		晴	雨	
X	晴	0.45	0.12	0.57
	雨	0.15	0.28	0.43
$P(y)$		0.60	0.40	

Mutual Information(2)

- Example
  - X: Real weather
  - Y: Weather forecast
  - $P(x,y)$: Individual probability
  - $P(x)$, $P(y)$: Joint probability

$P(x,y)$		Y: fine	Y: rain	$P(x)$
X	fine	0.45	0.12	0.57
	rain	0.15	0.28	0.43
$P(y)$		0.60	0.40	

相互情報量(3)

- 実際の天気のエントロピー: $H(X)$

$$H(X) = H_f(0.57) = 0.986$$

- ここに、H_f はエントロピー関数

$$H_f(p) \equiv -p\log_2 p - (1-p)\log_2(1-p)$$

- 天気予報を既知としたときの実際の天気の条件付確率

$$P(x|y) = \frac{P(x,y)}{P(y)}$$

Mutual Information(3)

- Entropy of real weather: $H(X)$

$$H(X) = H_f(0.57) = 0.986$$

- where、H_f denotes entropy function

$$H_f(p) \equiv -p\log_2 p - (1-p)\log_2(1-p)$$

- conditional probability of real weather when we know weather forecast

$$P(x \mid y) = \frac{P(x,y)}{P(y)}$$

相互情報量(4)

- 条件付確率(天気予報既知)
  - 天気予報が晴のときに実際の天気が晴、雨の確率は0.75, 0.25

$P(x\|y)$		Y	
		晴	雨
X	晴	0.75	0.30
	雨	0.25	0.70

- 天気予報"晴"が既知の場合の実際の天気のエントロピー: $H(X|f)$

$$H(X\,|\,f) = H_f(0.75) = 0.81$$

- 天気予報"雨"が既知の場合の実際の天気のエントロピー: $H(X|r)$

$$H(X\,|\,r) = H_f(0.70) = 0.88$$

Mutual Information(4)

- Conditional probability（under known weather forecast）
  - P(fine)=0.75, P(rain)=0.25 when weather forecast says "fine"

$P(x\|y)$		Y: fain	Y: rain
X	fine	0.75	0.30
	rain	0.25	0.70

- Real weather's entropy : $H(X|f)$ when forecast "fine" is known

$$H(X \mid f) = H_f(0.75) = 0.81$$

- Real weather's entropy : $H(X|r)$ when forecast "rain" is known

$$H(X \mid r) = H_f(0.70) = 0.88$$

相互情報量(5)

- 条件付エントロピー
  - 天気予報を既知としたときの実際の天気のエントロピー: $H(X|Y)$

$$H(X|Y) = 0.60 \times 0.81 + 0.40 \times 0.88 = 0.838$$

  - 条件付エントロピーの定義

$$H(X|Y) = -\sum_{y} P(y) \sum_{x} P(x|y) \log_2 P(x|y)$$
$$= -\sum_{x}\sum_{y} P(x,y) \log_2 P(x|y)$$

Mutual Information(5)

- Conditional entropy
  - Real weather's entropy with known weather forecast: $H(X|Y)$

$$H(X\,|\,Y) = 0.60 \times 0.81 + 0.40 \times 0.88 = 0.838$$

  - Definition of conditional entropy

$$H(X\,|\,Y) = -\sum_{y} P(y) \sum_{x} P(x\,|\,y) \log_2 P(x\,|\,y)$$
$$= -\sum_{x}\sum_{y} P(x,y) \log_2 P(x\,|\,y)$$

相互情報量(6)

- 相互情報量 (Mutual Information (content))

$$I(X;Y) \equiv H(X) - H(X|Y)$$

- 相互情報量は、情報によって減少したあいまいさの尺度
  - 天気予報の例では、$H(X)=0.986$, $H(X|Y)=0.838$ であり、相互情報量は $I(X;Y)=0.986-0.838=0.146$ となる
  - 天気予報によって、実際の天気に関して、平均 0.146 ビットの情報量が与えられることを意味する

Mutual Information(6)

- Mutual Information (content)

$$I(X;Y) \equiv H(X) - H(X|Y)$$

- Mutual Information is a measure of ambiguity by receiving information
  - In the example, $H(X)=0.986$, $H(X|Y)=0.838$. Thus, mutual information is given by $I(X;Y)=0.986-0.838=0.146$
  - This means that 0.146 bits information is given in average by weather forecast in regard to real weather