

# コンピュータ科学II

担当：武田敦志 <takeda@cs.tohoku-gakuin.ac.jp>  
http://takeda.cs.tohoku-gakuin.ac.jp/

## 今日の話

### ■情報理論

情報とは？

不確定な事象を確定させるもの

情報量とエントロピー

エントロピー：事象の不確定さ

情報を得ることによりエントロピーは減少する

減少したエントロピー = 情報利得

単位は bit

冗長量

表現の情報量と本当の情報量の差分

データの圧縮技術：冗長量を少なくする技術

page 2

## 情報(1)

### ■データ・情報・知識

#### ●データ (Data)

事実 (fact) を**表現**するもの

#### ●情報 (Information)

不確定な事象を**確定**するもの

⇒ データを整理することで得られる

#### ●知識 (Knowledge)

不確定な問題を**解決**するもの

⇒ 情報を整理することで得られる

⇒ 『知識+データ』で問題を解決する

page 3

## 情報(2)

### ■データ

2010年5月31日・仙台の気候データ

時間	気温(°C)	降水量(mm)	風向(16方位)	風速(m/s)	日照時間(分)
6時	10.9	0.0	北北西	2	60
7時	12.3	0.0	北	2	60
8時	13.8	0.0	北	1	60
9時	16.4	0.0	東	1	60
10時	17.0	0.0	南東	4	60
11時	16.9	0.0	南東	4	60
12時	17.9	0.0	東南東	4	60
13時	16.9	0.0	東南東	4	60
14時	16.9	0.0	東南東	4	60
15時	16.4	0.0	東南東	5	60
16時	16.4	0.0	南東	4	60
17時	15.0	0.0	南東	4	60
18時	13.8	0.0	南東	4	60

page 4

## 情報(3)

### ■情報

2010年5月31日の情報

最高気温：17.9 [°C]  
降水量：0 [mm]  
最高風速：5 [m/s]  
日照時間：60 [min/h]

5月31日はどの様な日？  
(不確定)

情報 → ↓

(確定)  
5月31日は  
暖かい・穏やか・快晴

2010年5月31日のデータ

時間	気温(°C)	降水量(mm)	風向(16方位)	風速(m/s)	日照時間(分)
6時	10.9	0.0	北北西	2	60
7時	12.3	0.0	北	2	60
8時	13.8	0.0	北	1	60
9時	16.4	0.0	東	1	60
10時	17.0	0.0	南東	4	60
11時	16.9	0.0	南東	4	60
12時	17.9	0.0	東南東	4	60
13時	16.9	0.0	東南東	4	60
14時	16.9	0.0	東南東	4	60
15時	16.4	0.0	東南東	5	60
16時	16.4	0.0	南東	4	60
17時	15.0	0.0	南東	4	60
18時	13.8	0.0	南東	4	60

page 5

## 情報(4)

### ■知識

気候に関する知識

天候が同様ならば  
翌日の気候も同様

気候に関する天候データ

5月31日は晴れ  
6月1日は晴れ

6月1日の気候は  
5月31日と同様になる

2010年5月31日のデータ

時間	気温(°C)	降水量(mm)	風向(16方位)	風速(m/s)	日照時間(分)
8時	13.8	0.0	北	1	60
9時	16.4	0.0	東	1	60
10時	17.0	0.0	南東	4	60
11時	16.9	0.0	南東	4	60
12時	17.9	0.0	東南東	4	60
13時	16.9	0.0	東南東	4	60
14時	16.9	0.0	東南東	4	60

2010年6月1日のデータ

時間	気温(°C)	降水量(mm)	風向(16方位)	風速(m/s)	日照時間(分)
8時	13.9	0.0	南東	1	60
9時	15.7	0.0	南東	2	60
10時	17.2	0.0	南東	3	60
11時	17.0	0.0	南南東	4	60
12時	17.4	0.0	南東	4	60
13時	16.9	0.0	南東	5	60
14時	17.8	0.0	南南東	6	60

page 6

## 情報(5)

### ■情報と不確定の事象

#### ●情報 (Information)

不確定な事象を**確定**するもの  
完全に確定するとは限らない  
⇒ **不確定の度合い**を減らす

(例) 8月1日の仙台の天気は？

晴 or 曇 or 雨 or 雪 ← 不確定の度合い：4択



情報：8月1日に仙台で雪は降らない

晴 or 曇 or 雨 ← 不確定の度合い：3択

page 7

## 情報量とエントロピー(1)

### ■不確定の度合い (エントロピー)

直感的には『予想不可能＝不確定の度合いが大きい』

例えば、

- 6面サイコロを投げる  
1～6のどの面が出るかは完全ランダム
- 宝くじを買う  
ほとんど外れ。たまに当たる。

不確定の度合い = **エントロピー** という

page 8

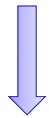
## 情報量とエントロピー(2)

### ■天気予報という情報

8月1日の仙台の天気は？

晴：確率 1/2  
曇：確率 1/4  
雨：確率 1/4

『晴』可能性が高いが  
『曇』『雨』かもしれない



情報：7月31日の天気予報  
『8月1日は晴れ。降水確率0%』

8月1日の仙台の天気は？

晴：確率 3/4  
曇：確率 3/16  
雨：確率 1/16

ほぼ確実に『晴』  
『雨』の可能性は非常に少ない

page 9

## 情報量とエントロピー(3)

### ■情報量

情報の量は数値で表現できる

⇒ 単位は **bit**

ある事象が存在し、その発生確率が  $P$  ならば  
この事象が発生することの情報量  $I$  は

$$I = -\log_2 P$$

となる

例：明日の天気予報の場合

晴：確率 1/2      『晴』という予報の情報量：

曇：確率 1/4      『雨』という予報の情報量：

雨：確率 1/4

page 10

## 情報量とエントロピー(4)

### ■平均情報量

情報量の高い情報は『めったに手に入らない』

⇒ 通常は**平均情報量**で考える

$n$  個の事象が存在し、それぞれの発生確率が

$$P_1, P_2, P_3, \dots, P_n$$

とすると、この平均情報量  $I$  は

$$I = -\sum_1^n P_i \log_2 P_i$$

となる

page 11

## 情報量とエントロピー(5)

### ■エントロピー（不確定の度合い）

エントロピーも数値で表現できる

情報によってエントロピー減少する

$i$  [bit] の情報で確定される =  $i$  [bit] のエントロピー

$n$  個の事象が存在し、それぞれの発生確率が

$$P_1, P_2, P_3, \dots, P_n$$

とすると、このエントロピー  $E$  は

$$E = -\sum_1^n P_i \log_2 P_i$$

となる

page 12

## 情報量とエントロピー(6)

### ■エントロピーの計算

例：明日の天気予報の場合

晴：確率 1/2

曇：確率 1/4

雨：確率 1/4

$$\begin{aligned} E &= -\sum_1^n P_i \log_2 P_i \\ &= -0.5 \log_2 0.5 - 0.25 \log_2 0.25 - 0.25 \log_2 0.25 \\ &= 1.5 \text{ [bit]} \end{aligned}$$

明日の天気を確定するためには 1.5 [bit] の情報が必要

page 13

## 演習問題(1)

### ■次の事象のエントロピーを計算せよ

6面サイコロを1回振ったときに出る目の数

( $\log_2 6 = 2.6$  として計算すること)

8本の「くじ」から1本を引く時のエントロピー  
ただし、入っている「くじ」は以下の8本とする

1本：A賞, 1本：B賞,  
2本：C賞, 4本：はずれ

page 14

## 情報利得の計算(1)

### ■エントロピーと情報利得の関係

情報取得前のエントロピー：E

情報取得後のエントロピー：E'

平均情報利得：I

$$I = E - E'$$

情報利得：

事象を確定した程度

⇒ 情報取得前と取得後のエントロピーの差分

page 15

## 情報利得の計算(2)

### ■取得した情報量

例：『それなりに当たる』天気予報

予報は「晴」か「雨」  
のどちらか

事象：晴：確率 1/2

曇：確率 1/4

雨：確率 1/4

予報：晴：確率 1/2

曇：確率 0

雨：確率 1/2



結果：予報が『晴』のとき

晴：確率 1

曇：確率 0

雨：確率 0

予報が「晴」なら  
確実に晴れる

予報が『雨』のとき

晴：確率 0

曇：確率 1/2

雨：確率 1/2

予報が「雨」の場合、  
半分の確率で曇りとなる

page 16

## 情報利得の計算 (3)

### ■取得した情報量

例：『それなりに当たる』天気予報の場合

事象のエントロピー

$$E = 1.5 \text{ [bit]}$$

予報が『晴』のときのエントロピー

$$E_f = -1.0 \log_2 1.0 = 0 \text{ [bit]}$$

予報が『雨』のときのエントロピー

$$E_r = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1 \text{ [bit]}$$

予報を聞いた後のエントロピー

$$E' = 0.5E_f + 0.5E_r = 0.5 \times 0 + 0.5 \times 1 = 0.5 \text{ [bit]}$$

天気予報の情報量

$$I = E - E' = 1.0 \text{ [bit]}$$

page 17

## 情報利得の計算 (4)

### ■取得した情報量

例：『まったく当たらない』天気予報

予報は「晴」か「雨」のどちらか

事象：晴：確率 1/2  
曇：確率 1/4  
雨：確率 1/4

予報：晴：確率 1/2  
曇：確率 0  
雨：確率 1/2



結果：予報が『晴』のとき

晴：確率 0  
曇：確率 1/2  
雨：確率 1/2

予報が「晴」なら曇りか雨になる

予報が『雨』のとき

晴：確率 1  
曇：確率 0  
雨：確率 0

予報が「雨」なら確実に晴れる

page 18

## 情報利得の計算 (5)

### ■取得した情報量

例：『全く当たらない』天気予報の場合

事象のエントロピー

$$E = 1.5 \text{ [bit]}$$

予報が『晴』のときのエントロピー

$$E_f = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1 \text{ [bit]}$$

予報が『雨』のときのエントロピー

$$E_r = -1.0 \log_2 1.0 = 0 \text{ [bit]}$$

予報を聞いた後のエントロピー

$$E' = 0.5E_f + 0.5E_r = 0.5 \times 1 + 0.5 \times 0 = 0.5 \text{ [bit]}$$

天気予報の情報量

$$I = E - E' = 1.0 \text{ [bit]}$$

page 19

## 冗長量(1)

### ■表現上の情報量と本当の情報量

表現上の情報量

情報を記録・送信するために必要な情報量

⇒ ファイルの大きさなど

本当の情報量

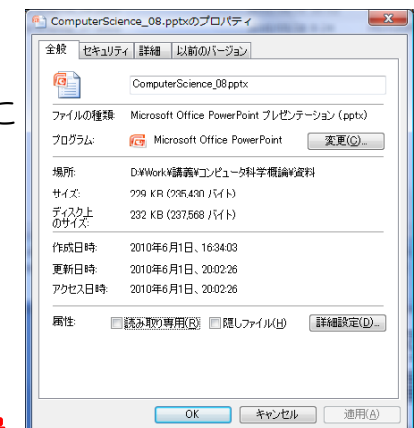
情報を表現するために必要な情報量

⇒ 事象のエントロピー

**表現上の情報量 ≥ 本当の情報量**

理論上は『本当の情報量』の長さで表現できる

表現上の情報量と本当の情報量の差分：**冗長量**



page 20

## 冗長量(2)

### ■表現上の情報量と本当の情報量

事象：晴：確率 1/2  
曇：確率 1/4  
雨：確率 1/4

情報：天気は晴れです

表現上の情報量

$$7 [\text{文字}] \times 24 [\text{bit/文字}] = 148 [\text{bit}]$$

本当の情報量

$$\text{エントロピー} = 1.5 [\text{bit}]$$

無駄な情報量

$$\text{冗長量} = 148 - 1.5 = 146.5 [\text{bit}]$$

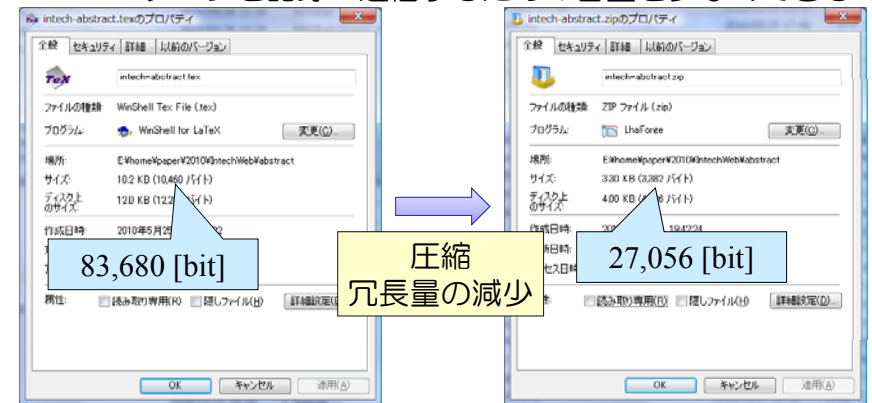
page 21

## 冗長量(3)

### ■冗長量と圧縮技術

冗長量を少なくすると、表現のための情報量が減る

⇒ データを記録・送信するための容量を少なくできる



page 22

## まとめ

### ■今日のまとめ

情報とは？

不確定な事象を**確定**させるもの

情報量とエントロピー

エントロピー：事象の不確定さ

情報を得ることによりエントロピーは減少する

減少したエントロピー = 情報利得

単位は **bit**

冗長量

表現の情報量と本当の情報量の差分

情報の圧縮技術：冗長量を少なくする技術

page 23

## 演習問題(2)

1年のうちの1日を選ぶときのエントロピーを求めよ  
ただし、小数点は繰り上げること

月と日を、それぞれ 5bit で表現した場合の冗長度を求めよ  
ただし、小数点は切り捨てること

表現の例 6月12日： 00110 / 01010

12月31日： 01010 / 11111

page 24