情報理論
第 1 回 イントロダクション

堀田 政二
工学部 情報工学科

今回のアウトライン

- 情報理論を学ぶ意義
- 情報理論における通信路モデル
- シラバス，講義の概要 · 計画
- デジタルの復習
- 2 進数の復習

情報とは何か

- **情報** (information) **の語源**
  - **英語の** information: (**心において**) form (**形**) **を与える，転じて「あるものごとの内容や事情についての知らせ」**
  - **日本語の情報**: **「敵情の報告」 を縮めたもの** "**仏國歩兵陣中要務實地演習軌典**"**，酒井忠恕訳，**1876
- **情報の特徴**
  - **そもそも受け取る，伝えるという意味だが，コンテンツ，知識という意味でも使われる**
  - **ただし，知識は受け取る側にとって新鮮でなければならない**
  - **しかし新鮮かどうかは人間** (**受け取り側**) **に依存する**

主観的で捉えどころのない "**情報**" **を計測可能なもので定義したい**

情報理論で扱う情報とは

確率を使って表現可能なあらゆる知識

【例題】

1. 100 本のクジがあり，1 本がアタリ，99 本がハズレ
2. 甲と乙が順番にクジを引く
3. まず，甲が 100 本のクジから 1 本を引き，内容を確認せずに保持する
4. 次に，乙が 99 本のクジから 1 本を引き，内容を確認したところハズレであった
5. この時点で甲のクジがアタリである確率はいくらか？

(4)

確率は変化する

- ③の時点では甲がアタリである確率は $1/100$ (事前確率)
- ところが④の事実 (知識) を知ったときに，甲がアタリである確率は $1/99$ に増大する (事後確率)

乙はハズレ，という知識 (情報) を受け取ることで，甲がアタリである不確定度 (曖昧さ) は減少した $\Rightarrow$ この減少分を情報量 (information content) と定義

- もし乙のクジがアタリであった場合は，甲がアタリである確率は 0

情報理論の根幹にある考え方

稀にしか起こりえない現象が起こったことを知った時に得られる情報量は大きい

(5)

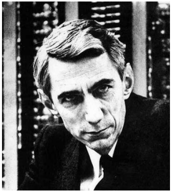
情報理論の考え方はなぜ重要か

- 情報を確率の視点から捉えることで，情報に関する人間の解釈を排除できる
- そのため「情報」が学問として体系づけられ，科学的・工学的な取り組み対象となった
- この結果，インターネットに代表される通信，符号化，画像や動画の圧縮技術が発展することとなる

情報理論を学ぶ意義

今日の情報社会の基礎となる理論であり，その考え方は情報工学に関わる他の科目とも相通じるものが多い．そのため，情報理論は本学科で必須科目 (2 単位) となっている

クロード・エルウッド・シャノン



Claude Elwood Shannon
1916 年 4 月 30 日-2001 年 2 月 24 日

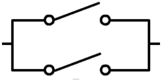
アメリカの電気工学者，数学者．情報理論の父

情報理論の創始的な論文 C. Shannon, “A Mathematical Theory of Communication”, Bell System Technical Journal, vol. 27, pp. 379–423, 623–656, 1948

スイッチのオン・オフと，命題の真・偽の関係に対応させると並列接続は OR**，直列接続は** AND **であることを示し，論理演算がスイッチ回路で実行できることを証明**

C. Shannon, “A Symbolic Analysis of Relay and Switching Circuits”, MIT, Dept. of Electrical Engineering, master’s thesis, 1938

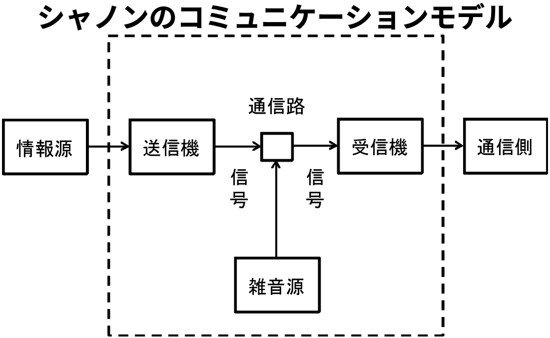
コンピュータを論理演算機として利用可能に



並列

直列

情報の伝達，コミュニケーションモデル



- 情報源から発せられたメッセージは送信機で符号化され，信号として通信路に流される
- 通信路 (通信回路等) を通った信号は受信機にて復号化された後，通信側に提示される
- 情報理論で扱うのは点線部分のみ

(8)

シャノンの功績

- 情報理論は点線部分，すなわち通信路を流れる信号のみ注目する
- これは工学の使命である，与えられた課題をより速く，より正確に，より安く解決するために，情報を工学的に扱えるようにするためである
- シャノンの功績を簡単にまとめると，以下のようなものになる
  - 情報量の考案
  - 通信路の性能を限界まで発揮できる符号化理論の構築 (情報源符号化定理，またはシャノンの第一基本定理)
  - 通信路にノイズが加わっても，元の信号をいくらでも小さな誤り確率で伝送できる符号化法の存在を証明したこと (通信路符号化定理，またはシャノンの第二基本定理)

情報源符号化 (source coding)

情報源符号化

情報源を効率のよい符号に変換すること

送信したい情報を 1 と 0 の組合せで表すことを考える
例えば，はい，いいえ，どちらでもない，を送るとする

情報	発生確率	符号 C_1	符号 C_2
はい	0.3	0	1
いいえ	0.2	1	10
どちらでもない	0.5	10	0

C_1 より C_2 の方が効率が良い

通信路符号化 (channel coding)

通信路符号化

通信路で生じるノイズに起因する通信誤りが小さくなる符号化

例えば，入力と同じ出力になる確率が 0.9，異なる出力となる確率が 0.1 となる通信路を考える

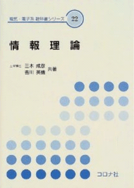
情報	符号 C_1	符号 C_2	誤り確率
はい	1	111	0.028
いいえ	0	000	0.028

- 符号の長さを 3 倍にして，多数決で 0 か 1 かを決定する
- C_1 より C_2 の方が誤りが小さくなる．その代り 3 倍の送信時間が必要

2012 年度 授業予定

1. 4/09 **イントロダクション，シラバス，二元符号と二進数**
2. 4/16 **確率論の基礎**
3. 4/23 **確率分布**
4. 4/30 **確率過程**
5. 5/07 **情報量の定義**
6. 5/14 **情報源モデルと通信路モデル**
7. 5/21 **中間試験**
8. 5/28 **中間試験の解説& 相互情報量**
9. 6/04 **伝送情報量と通信路容量**
10. 6/11 **休講**
11. 6/18 **情報源符号化の基礎**
12. 6/25 **通信路符号化の基礎と情報源符号化**
13. 7/02 **拡大情報源と動的符号化**
14. 7/09 **パリティ検査とハミング符号，線型符号と巡回符号**
15. 7/16 **期末試験**

参考文献と評価方法



三木，吉川，“情報理論”，コロナ社，2000

- 塩野，“わかりやすいディジタル情報理論”，オーム社，1998
- 小川，“マルチメディア時代の情報理論”，コロナ社，2000
- 大石，“例にもとづく情報理論入門”，講談社サイエンティフィック，1993
- 神谷，川島，“情報・符号理論 — デジタル通信の基礎を学ぶ —”，オーム社，2012

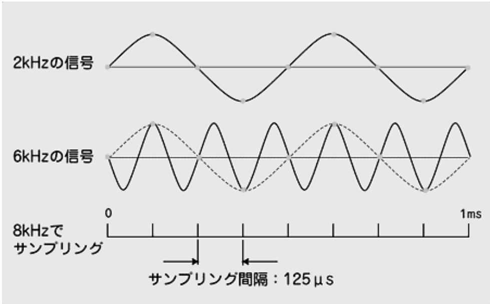
成績評価の方法

出席状況 10%，中間試験 40%，期末試験 50%

アナログ/デジタル変換 (A/D 変換) における標本化

標本化定理 (シャノン・染谷の定理)

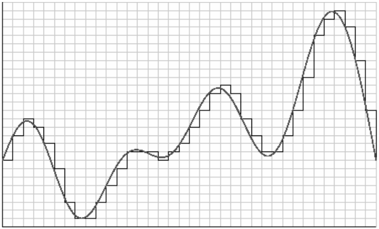
入力信号の最高周波数の 2 倍以上の周期で出力値を取り出せば（サンプリングすれば），元の入力信号を復元可能



アナログ/デジタル変換における量子化

量子化 (quantization)

サンプリングされた連続値を，最も近い離散値で置き換えること



- 離散値の種類数と量子化誤差はトレードオフとなる

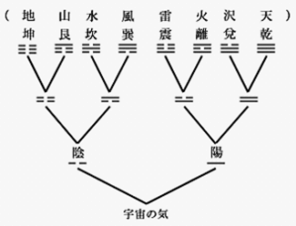
デジタル化とは

時間や空間をサンプリングし，サンプリングされた値を量子化することで実現

- 音声：時間をサンプリング，振幅値を量子化
- 画像：空間をサンプリング，明るさ (ピクセル値) を量子化

二進記数法 (二進数) によるデジタル化信号の符号化

- デジタル化された値は，0 と 1 の組合せ (二進記数法) で符号化される
- 最小単位 (一桁の二進数) をビット (bit) と呼ぶ
- N ビットで表せる数は 0 から $2^N - 1$ まで
    - $N = 3$ ビット：0 から $2^3 - 1 = 7$ までの 8 種類
    - $N = 4$ ビット：0 から $2^4 - 1 = 15$ までの 16 種類
    - $N = 5$ ビット：0 から $2^5 - 1 = 31$ までの 32 種類
- 例：デジタル体温計が $32.0°$C から $42.0°$C まで $0.1°$C 刻で測定可能な場合，体温は 101 通りとなる．これを表現するためには 7 bit (128 通り) が必要となる



(16)

文字コード

計算機では文字に番号 (文字コード) を割り当てて表現する

- **英語圏では** ASCII (American Standard Code for Information Interchange)
  - **数字：**0 **から** 9 **の** 10 **種類**
  - **アルファベット：**A **から** Z **の大文字，小文字，合わせて** 52 **種類**
  - **その他記号：**$+$**，**$-$**，**$\%$**，括弧等，数十種類**
  - **結局** 7bit **で十分**
- **日本語では** JIS **コード，**EUC **コード，**Shift-JIS **など**
  - **日本語ではさらに漢字，平仮名，カタカナが加わる**
  - **結局** 16 bit (2 byte$=2^{16}=65536$ **種類**) **が必要**
  - **新聞** 1 **ページは** 256K bit **程度らしい**
  - 8 bit= 1 byte (B), 1KB=1024 B, 1MB= 1024KB

ASCII コードの実際

000:	013:♪	026:→	039:'	052:4	065:A	078:N	091:[	104:h	117:u
001:☺	014:♫	027:←	040:(	053:5	066:B	079:O	092:\	105:i	118:v
002:☻	015:☼	028:∟	041:)	054:6	067:C	080:P	093:]	106:j	119:w
003:♥	016:►	029:↔	042:*	055:7	068:D	081:Q	094:^	107:k	120:x
004:♦	017:◄	030:▲	043:+	056:8	069:E	082:R	095:_	108:l	121:y
005:♣	018:↕	031:▼	044:,	057:9	070:F	083:S	096:`	109:m	122:z
006:♠	019:‼	032:	045:-	058::	071:G	084:T	097:a	110:n	123:{
007:•	020:¶	033:!	046:.	059:;	072:H	085:U	098:b	111:o	124:¦
008:◘	021:§	034:"	047:/	060:<	073:I	086:V	099:c	112:p	125:}
009:○	022:▬	035:#	048:0	061:=	074:J	087:W	100:d	113:q	126:~
010:◙	023:↨	036:$	049:1	062:>	075:K	088:X	101:e	114:r	127:⌂
011:♂	024:↑	037:%	050:2	063:?	076:L	089:Y	102:f	115:s	
012:♀	025:↓	038:&	051:3	064:@	077:M	090:Z	103:g	116:t	

10 進数から 2 進数への変換 (整数部分)

- 10 進数の 123.159 を 2 進数で表すことを考える
- まずは整数部分の 123 を 2 進数で表現する
- 2 で割った商と余りを下から順に左から並べる

$$
\begin{array}{r@{\,}l@{\quad}c@{\quad}c}
2 & \underline{)\ 123} & & \\
2 & \underline{)\ 61} & \cdots & 1 \\
2 & \underline{)\ 30} & \cdots & 1 \\
2 & \underline{)\ 15} & \cdots & 0 \\
2 & \underline{)\ 7} & \cdots & 1 \\
2 & \underline{)\ 3} & \cdots & 1 \\
 & \ \ 1 & \cdots & 1
\end{array}
$$

- $(123)_{10} \to (1111011)_2$

10 進数から 2 進数への変換 (小数点部分)

- 10 進数の小数点部分に 2 を掛ける
- 得られた値が 1 以上になったら 1 を引き，2 を掛ける．1 未満であればそのまま 2 を掛ける
- 上記を所望の桁まで計算し，小数点を表すピリオドの左の 0 と 1 を上から順に左から並べる

$$
\begin{aligned}
0.159 \times 2 &= 0.318 \\
0.318 \times 2 &= 0.636 \\
0.636 \times 2 &= 1.272 \\
\underline{0.272} \times 2 &= 0.544 && \text{1 を引く} \\
0.544 \times 2 &= 1.088 \\
\underline{0.088} \times 2 &= 0.176 && \text{1 を引く} \\
0.176 \times 2 &= 0.352
\end{aligned}
$$

- $(0.159)_{10} \to (0010100)_2$
- したがって $(123.159)_{10} = (1111011.0010100)_2$

2 進数から 10 進数への変換

- 2 進数の i 番目の桁を a_i (0 か 1) とする．添え字 i が 0 以上の部分を整数部分，i が負の部分を小数点部分とする

$$a_N a_{N-1} \cdots a_1 a_0 . a_{-1} a_{-2} \cdots$$

- 10 進数に変換するには以下を計算すればよい

$$a_N 2^N + a_{N-1} 2^{N-1} + \cdots + a_1 2^1 + a_0 2^0 + \frac{a_{-1}}{2} + \frac{a_{-2}}{2^2} \cdots$$

- $(1111011.00110)_2$ は
  - 整数部分：$64 + 32 + 16 + 8 + 2 + 1 = 123$
  - 小数部分：$1/8 + 1/32 = 0.156...$
- したがって $(1111011.00110)_2 = (123.156...)_{10}$

問題

1. **次の 10 進数を 2 進数に変換せよ**
   (1)365　(2)21.5
2. **次の 2 進数を 10 進数に変換せよ**
   (1)101011　(2)101.11
3. $0 \leq x(t) \leq 15$ **の値を出力する連続関数** $x(t)$ **がある．この関数を** $t = 0.1$ **の間隔でサンプリングしたところ，以下の 5 つの値が観測された．それぞれの値を量子化し，4bit で符号化せよ**

$x(t)$	量子化	符号化
6.2		
3.1		
9.4		
11.3		