

ICT リテラシー (情報技術論) A

-- 第 09 回 : 情報量 --

栗野 俊一

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く
禁じます

2023/11/07 ICT リテラシー (情報技術論) A

伝言

私語は慎むように !!

□ 席は自由です

- できるだけ前に詰めよう
- コロナ対策のために、ソーシャルディスタンスをたもとう

□ 色々なお知らせについて

- 栗野の Web Page に注意する事

<http://edu-gw2.math.cst.nihon-u.ac.jp/~kurino>

- google で「kurino」で検索

前回 (第 08 回) の復習

ICT リテラシー (情報技術論) A

前回 (第 08 回) の復習

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

前回 (第 08 回) の復習

□ 前回 (第 08 回) の復習 (メディアリテラシー [Mail / SNS])

○ 一般のメディアリテラシー (発信者)

- ▶ 個人情報の扱い (画像 / メディア固有 [Mail Address / Line ID])
- ▶ 公序良俗への配慮 (他人に不快感を与える内容/表現をさける)
- ▶ トラブルの回避 (対面[会話]との違い => 誤解の発生 / 冷静な対応)

○ メールメディアリテラシー

- ▶ 昔からあるメディア : 通信相手への配慮 (作法/形式/署名)
- ▶ メールアドレスの扱いへの配慮 (Bcc の扱い)

○ SNS 共通のメディアリテラシー

- ▶ SNS の拡散機能への配慮 (何に加担しているのか? / 程度の問題)
- ▶ 設定の確認 : 契約内容やデフォルト設定の確認

○ SNS メディア固有のリテラシー

- ▶ Twitter : リツイート (RT) は慎重に
- ▶ Facebook : 公開範囲の設定 / 「いいね!」や「シェア」の抑制
- ▶ Line : 既読の扱い (飽く迄も安否確認手段)

今週 (第 09 回) の概要

ICT リテラシー (情報技術論) A

今週 (第 09 回) の概要

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

今週 (第 09 回) の予定

□ 今週 (第 09 回) の予定

○ 講義：情報量

- ▶ 情報量の意味 (Text p.61, 5.2 節)
- ▶ 情報量の計量 (Text p.62, 5.3 節 - p.67, 5.9 節)
- ▶ 情報圧縮技術 (Text p.68, 5.10 節 - p.68 5.11 節)
- ▶ 誤り検出/訂正技術 (Text p.71, 5.12 節)

今週 (第 09 回) の目標

□ 今週 (第 09 回) の目標

○ 講義

- ▶ アナログとデジタル(情報の符号化)の違いを知る
- ▶ 情報量の観点から、情報を扱う
- ▶ 情報量の計算が行えるようにする
- ▶ 情報圧縮や情報訂正等の技術の概要を把握する

今週 (第 09 回)

- 前回 (第 08 回) の課題
 - Web Class「小テスト-08」
- 今週 (第 09 回) の課題
 - Web Class「小テスト-09」

情報量

ICT リテラシー (情報技術論) A

情報量

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

情報と情報量

□「情報」とは

○「確率的に振る舞う事象」の「内の一つ」を「示した表現」

▶ 例1：『本日(2023/11/07)の午後の天気』は『雨』(降水確率：50%)

▶ 例2：『サイコロを振った』ら『3の倍数の目が出た』(確率は 1/3)

○「『情報』を得る」と(それが正しいならば..)「確率的な状況が『確定』」する

▶ 例3：歌手のコンサートは来月の末日 => 有給の予定が立てられる

○それ(情報が指す事象)が起きる確率が低い程、情報の価値は高い

▶ 「情報の量(E)」は、「確率(P)」と関係がある ($E = -\log_2 P$)

▶ 「確定した事実(確率=1)」は「情報量」がない[0] (そんなの「知って」た..)

□情報の表現(保存/伝達)

○「情報」に対応した、「物理状態[エネルギー]」に変換(表現)する

▶ cf. 通信の OSI 7 階層モデルの最下層は「物理層」になっている

▶ 「A」という文字：声で伝える場合 => 「エー」という「音声」で伝える

○「情報」をどのような形(表現形式[メディア]、対応規則)にするかは応用次第

▶ 物理現象の影響を受ける (その特性に従って、選択する)

アナログとデジタル

□ アナログ(連続量)とデジタル(離散量)

○ アナログ (連続量 : 実数値) : 物理量は、基本アナログ (表現に無限の長さ必要)

▶ 「時間」は、連続に流れる

○ デジタル (離散量 : 整数値) : 観測量は、基本デジタル (表現が有限の長さで可能)

▶ 「時刻」は、離散的しか表現できない (有限の表現形式から一つを選ぶ)

○ アナログからデジタルに変換すると「誤差」が生じる (精度落ち)

▶ 逆変換は常に可能

▶ 「測定(アナログからデジタルに変換)」時に、どこまでの精度を要求するか？

▶ 「誤差」が「気にならない」なら、デジタルにしても問題ない (1 秒を惜しむか？)

○ アナログ値には、(原理的に..) 無限の情報が埋め込める

▶ 拡大縮小が可能

○ 学問の差

▶ 数学は「無限(アナログ)」を扱う、情報(工学)は「有限(デジタル)」を扱う、物理(理学)は、「誤差(アナログとデジタルの差)」を扱う

□ 「デジタル」コンピュータ (論理回路)

○ 電圧の「高/低」で、符号の「1/0」を表現

▶ 「情報」を「『1/0』の『有限個』の組み合わせ」で表現(有限の表現) => デジタル

○ cf. 「アナログ」コンピュータもある

bit : 情報量の単位

□ 1 bit (Binary [2 の] Digit [数字])

○ (等しい確率の) 二つ (0/1) の内どちらかを表す情報の持つ情報量

▷ 例: コイントスで、表が出た/女の子が生まれた[厳密には違う]

○ 1 bit で「 $\{|0, 1|\} = 2 (= 2^1)$ 通り」の内の一つを表現できる

▷ 2 bit では、「 $\{|00, 01, 10, 11|\} = 4 = 2^2$ 通り」の表現が可能

▷ n bit では 2^n の表現が可能 ($0 \sim 2^n - 1$ に対応)

▷ 10 bit では $2^{10} = 1024$ '≒, $1000 = 10^3 = 1 \text{ K}$ の表現が可能

▷ $M = 10^6 = (10^3)^2$ '≒, $(2^{10})^2 = 2^{20} / G = 10^9 / T = 10^{12} / P = 10^{15} / ..$

□ 符号表現

○ 有限集合 (要素が有限 [M] 個の集まり) の要素の整数値表現

▷ $0 \sim M-1$ の整数値で対応

○ 2 進表現 : $M \leq 2^N$ となる N を考え、N bit で表現

○ 例: 曜日(の表現)の情報量 ($0 \sim 6$ の 7 通り)

▷ $2^2 = 4 < 7 < 8 = 2^3$ なので 2 bit より多く 3 bit より少ない情報量 ($\log_2(7)$ '≒, 2.8)

○ 例: キーボード上の文字 (英大文字/小文字、数字、記号、etc .. $7\text{bit} + 1\text{ bit} = 8\text{ bit} = 1\text{ byte}$)

色々なメディアの情報量

ICT リテラシー (情報技術論) A

色々なメディアの情報量

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

文字の表現と情報量

□ 文字の表現 (Text p.63, 5.5 節 ~ p.65, 5.6 節)

○ 「文字集合」を決める (要素と[有限な]要素数[M]が決る)

- ▶ 「文字集合」に整数値を割り当てる (0 ~ M-1)
- ▶ bit 表現に対応 ($2^{(N-1)} < M \leq 2^N$ となる N を選ぶ)

○ 文字表現の例

- ▶ ASCII : キーボード上の文字 (1 byte = 8 bit で表現)
- ▶ JIS Code (JIS:日本工業規格) : 日本語文字(漢字、ひらがな等)を定める (2 byte)
- ▶ Shift JIS Code (Microsoft) : JIS Code と半角カナの両方を含む (2 byte) Windows で利用
- ▶ EUC (AT&T) : Unix で利用される
- ▶ UTF-8 (ISO) : 世界中で利用されている文字を含む (3 byte)

○ 「1 文字」の情報量

- ▶ 「文字表現」によって異なる (ASCII => 1 byte / UTF-8 => 3 byte)
- ▶ 情報量 \Leftrightarrow それを表現するために必要な bit 数

○ 「文字化け」の問題

- ▶ コードが共通でないとおきる問題 (対応する文字がない / 解釈が異なる)

音声の情報量

□ 音声の情報量 (Text p.65, 5.7 節)

○ 音の物理量は、波(音波)で表現されるアナログ値 (各時刻の周波数)

▶ サンプル周波数：デジタル化をする時間間隔 (1 秒間に何度(n)、観測するか [n Hz])

▶ 量子化数：デジタル化をする波の高さの最大値 (byte 数)

○ 音声の情報量の例：サンプル周波数 => 44.1 kHz / 量子化数 => 2 byte

▶ 1秒あたりの情報量：44,100 x 2 byte =, 86 KB

□ 忠実性と効率のトレードオフ

○ サンプル周波数や量子化数を大きくすれば、より忠実になるが効率がわるくなる

静止画像の情報量

□ 静止画像の情報量 (Text p.66, 5.8 節)

○ デジタル画像 : (色のついた)点の矩形(縦横)の集まり

▶ 画像のデータの情報量 : 点の情報量 x 縦 x 横

▶ 点の情報量 : 白黒 => 1 bit / カラー => 3 原色 (+ 透明度) x 色調

○ 静止画の情報量の例 : カラー (3 原色 x 256 階調) / 画面サイズ (1920 x 1080)

▶ 1枚あたりの情報量 : $3 \times 1920 \times 1080 \text{ bit}$, 5.9 MB

□ 印刷物の情報量

○ 印刷物も、画像と同じように点の集まり

○ 解像度の単位 : dpi (dot per inch) 1 inch = 2.54cm 当りの dot 数

○ 印刷物の情報量の例 : 600 dpi のプリンタ / ハガキ (14.8 cm x 10 cm) 印刷

▶ 縦 : $600 \times 14.8 / 2.54 = 3496 \text{ dot}$

▶ 横 : $600 \times 10 / 2.54 = 2362 \text{ dot}$

▶ 全体 : $3496 \times 2362 \times 3 \text{ bit}$, 8.3 M ピクセル (830 万画素)

動画像の情報量

□ 動画像の情報量 (Text p.67, 5.9 節)

○ 動画 : 高速に静止画像を更新 (パラパラ漫画と同じ)

▶ 映画 : 1 秒間に 24 回の書き換え / ハイビジョン : 1 秒間に 60 回

○ 1 秒間の動画の情報量 : 静止画の情報量 \times 1 秒間の書き換え回数

○ 動画像の情報量の例 : ハイビジョン (1 秒間に 60 回) / 解像度 (3 x 1920 x 1080)

▶ $60 \times 5.9 \text{ MB} = 354 \text{ MB}$

通信の情報量

□ 通信の情報量 (Text p.68, 5.10 節)

○ 帯域 : 通信速度 / 通信周波数の範囲

- ▶ 通信周波数の範囲が広いと(並列して情報が送れるので)通信速度が速くなる
- ▶ 周波数が高いと、1秒間に送れる情報量が増える(単純に、通信速度が速くなる)

○ 通信速度の単位 : bps (Bit Par Second : 1秒間に送れる bit 数)

- ▶ ブロードバンド(広帯域) : 1 M bps 以上
- ▶ ナローバンド(狭帯域) : 1 M bps 以下

情報圧縮

ICT リテラシー (情報技術論) A

情報圧縮

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

情報圧縮

□ 情報圧縮 (Text p.68, 5.11 節)

○ 情報圧縮とは

- ▶ 対象となる情報の表現を変更する事により、より少ない量で表現する手段
- ▶ 元の情報を失わない「可逆圧縮」と、情報を失う「非可逆圧縮」がある

○ 一般の(符号)表現

- ▶ 全ての表現に対する事象が、同じ確率で生じる事を想定 => 現実と異なる
- ▶ 例：英語のアルファベット => 「e」が出る確率は $13\% > 3.4\% = 1/26$

○ ハフマンコーディング (可逆圧縮の例)

- ▶ 出現確率の高い文字を短い符号で表現すれば全体として短くなる
- ▶ zip/lzh/gz 等のファイルの圧縮に利用されている

○ 例: A:1/2, B:1/4, C,D:1/8 : ACAABBAD

- ▶ A,B,C,D : 00,01,10,11 => 00 10 00 00 01 01 00 11 (16 bit)
- ▶ A,B,C,D : 0,10,110,111 => 0 110 0 0 01 01 0 111 (14 bit)

○ ランレングス圧縮

- ▶ 同じ符号が連続して長く出る確率が高い：文字そのものより出現回数(>2)をかく
- ▶ 例: AAAABDDDDDDDC (12 byte) => A4BD6C (6 byte) [元のデータは数字はでないと仮定]

情報圧縮の例 (1)

□ 情報圧縮手段の違い

- 表現する情報の対象の性質(確率分布や範囲)によって、圧縮手段が異なる

□ (小さな数値の) ASCII 形式とバイナリ形式

- 小さな数値は、数字並び(ASCII 形式)より、その値に対応した数値(バイナリ形式)で表現した方がよい

▶ 例: "123" => 文字(ASCII)なら 3 byte = 24 bit / 数値ならば $127=2^7-1$ 以下なので 7 bit で済む

▶ ポイント: 「小さな数値の範囲」だから..

□ 音声の圧縮技術 (Text p.69)

- 音声: 色々な周波数の音波の集まり

▶ 人間の耳には、高周波が聞こえない => カットしても大丈夫

▶ (電話での)会話と(CDでの)音楽でも、圧縮技術が異なる

- MP3 の圧縮技術 (非可逆圧縮)

▶ 音声を周波数毎の音に分割(フーリエ変換) => 高周波をカット => 元に戻す(逆フーリエ変換): 結果的に情報量が減る

情報圧縮の例 (2)

□ 画像の圧縮技術 (Text p.69)

- 画像データの特徴：近接しているドットの情報は似ている(変化が小さい)

- ▶ 隣接したドットの変化量を考えると、小さい変化の起きる確率が高い

- ▶ 変化量の分布：周波数 => フーリエ変換が使える

- GIF/JPEG 形式：非可逆圧縮 (BMP は、非圧縮形式)

□ 動画の圧縮技術 (Text p.70)

- パラパラマンガ方式 => 連続した静止画は、変化が小さい

- ▶ 差分情報を利用：MPEG (Motion Picture Expert Group)

- ▶ MP3 (MPEG の Layer-3 : 音楽パートの圧縮技術)

□ 再現性と効率のトレード

- 非可逆圧縮の場合、圧縮率の高い方法とすると効率は高まるが再現率が落ちる

誤り検出

ICT リテラシー (情報技術論) A

誤り検出

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

誤り検出

□ 誤り検出 (Text p.71, 5.12 節)

○ 誤り(Error) : 情報の伝達の途中で、符号[情報を表す物理的な信号]の「異常」変化が起きる事がある

▶ 例: 音声での雑音 (S/N 比) / DVD の表面の汚れによる読取失敗

○ 誤りを検出し、小さな誤りなら訂正する技術が必要

□ 冗長性 : 「『同じ情報』に対する表現を『繰り返し与える』」というメタ表現

○ 例 : 「大事な(きちんと伝えたい)事だから、二度(繰り返し)言いました」

▶ 「聞き落とし[丁度教室が騒かった]」と言う「エラーを訂正する」仕組み

○ 係り受け : 拝啓 ~ 敬具(手紙) / こそ ~ 已然形(古語) / 主語(三単現)と動詞変形[s] (英語)

□ 誤り検出/訂正の原理

○ 符号表現に冗長性を与え、その冗長性(表現)が持つ情報を利用して、エラー検出/訂正を行う

誤り検出

□ パリティ (偶奇性)

- ASCII Code の文字は 7bit (128 種類) で表現可能

 - ▶ 1 bit のパリティbit を追加して、8 bit = 1 byte で通信

- パリティbit : 文字コードの 1 の出現個数は、偶数か奇数

 - ▶ パリティ bit の 0/1 を加えて、全体の 1 の出現個数の偶奇数をそろえる

 - ▶ 奇数パリティ: 全体の個数を奇数にする

 - ▶ 偶数パリティ: 全体の個数を偶数にする

- パリティbit を利用したエラー検出

 - ▶ 1 文字 : 1 byte = 8 bit 中で、1 bit のエラー (0/1 -> 1/0) がおきると、パリティが変化するので、エラー検出可能(訂正はできない)

□ CRC (Cyclic Redundancy Check : 巡回冗長検査)

- 検査 bit (冗長性を持つ bit) を増やし、工夫する事により、訂正も可能にした符号

□ 安全性と効率のトレード

- 冗長性を高める事により、安全性は高まるが、符号効率は劣化する

おしまい

ICT リテラシー (情報技術論) A

おしまい

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます