



第1回 情報理論とは

工学部 情報エレクトロニクス学科

情報科学研究所 情報理工学部門
大規模知識処理研究室

堀山 貴史



教科書の構成

- 中村、喜田、湊、廣瀬 著

基礎から学ぶ 情報理論 (第2版)、ムイスリ出版

- | | | |
|-------------------------------|---|-------------------|
| 1. 情報理論とは | } | 概要と準備 |
| 2. 情報量 と エントロピー | | |
| 3. 情報源のモデル | } | 情報源 とその符号化 |
| 4. 情報源符号化の限界 | | |
| 5. 情報源符号化法 | | |
| 6. 通信路のモデル | } | 通信路 とその符号化 |
| 7. 通信路符号化の限界 | | |
| 8. 通信路符号化法 | | |



情報

■ そもそも情報ってなに？

[株式会社岩波書店 広辞苑第五版]

- ① あることがらについての**しらせ**。「極秘—」
- ② 判断を下したり行動を起したりするために必要な、種々の媒体を介しての**知識**。「—が不足している」

■ もうちょっと分かるように言えないかなあ？

- 伝達されうる知識 ← 「情報」って何か知っている人にしか分からないのでは？
- 片方の(脳内)世界に存在する事象・概念を、他方の世界へ伝えるもの ← ???

とにかく、何か伝わるもので、物質とは性質が違う



情報と物質の違い

- たとえば、新聞紙そのものは物質に過ぎない。(＝情報の媒体)
 - － 読み終わった古新聞はただの紙。
 - － 新聞に書いてある内容に意味がある。
- 「物」は誰に送っても同じ「物」である。
- 情報は受け手によって大きく違う。
 - － 情報理論の教科書を小学生に読ませても、たぶん何も伝わらない。
 - － 大学生が読めば、小学生よりもずっと多くの情報が伝わる。
 - － 情報理論の専門家が読むと、すでに知っていることばかりで、大した情報は伝わらないかも知れない。

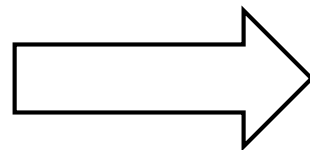
情報 = 受け手の知識が、どのくらい変化したか

これを数学的に定義して、学問として扱いたい → どうするか？



情報伝達の例

- 札幌で暮らしている人
- 冬の朝に起きて、今日の天気が雪と知っても、驚かない
 - いつも雪だから（受け手の知識は、あまり変わらない）
- 雨が降ったら、アレって思う
 - レアな事象だから（この情報で、受け手の知識は、大きく変わる）



今日の天気

晴、曇、雨、雪



受け手

天気の統計
(統計的知識)

晴	5.5 %
曇	1.2 %
雨	0.2 %
雪	93.1 %



情報理論で扱う数学的モデル

- 電信、電話、通信などで扱われる、文字列、音声、画像のようなデータの列であれば、普遍的・客観的に扱える
 - つまり、ある文字集合(アルファベット)の中から選ばれる記号が、一定時間ごとに記号の系列(データ)として流れてくる場合を考える

札幌の天気の場合だと、晴、曇、雨、雪が、時々刻々流れてくる

- この系列は、ある統計的な確率分布を持っているとする

札幌の天気の場合だと、晴、曇、雨、雪の天気の統計

受け手が元々知っている確率分布(知識)が、記号の系列(データ)を受け取ることによって変化するならば、その変化を「情報」と定義しよう

確率分布は数学的に記述でき、学問として扱いやすい
→ 理論として大きく発展し、さまざまな応用が広がった



情報理論の応用先

- 情報理論は様々な情報処理技術の基礎となる重要な理論
 - 情報エレクトロニクス学科の全コースの必修科目

(主な応用例)

- 電話、インターネット通信、無線通信、CD-ROM、DVD、ハードディスク、ラジオ・テレビ放送、天気予報、各種統計・予測、パターン認識、画像処理、データ検索、データマイニング、遺伝子情報処理、etc

(要素技術)

- 自動誤り訂正、データ圧縮、暗号化、情報量の概念、相互情報量、エントロピー、ひずみ、標本化



情報理論の応用例

データ量を1/16に削減。
でもほとんど問題ない

画像圧縮

BMPファイル
(.bmp; 295 MB)

ひずみを許す
符号化法へ変換



JPEGファイル
(.jpg; 18 MB)

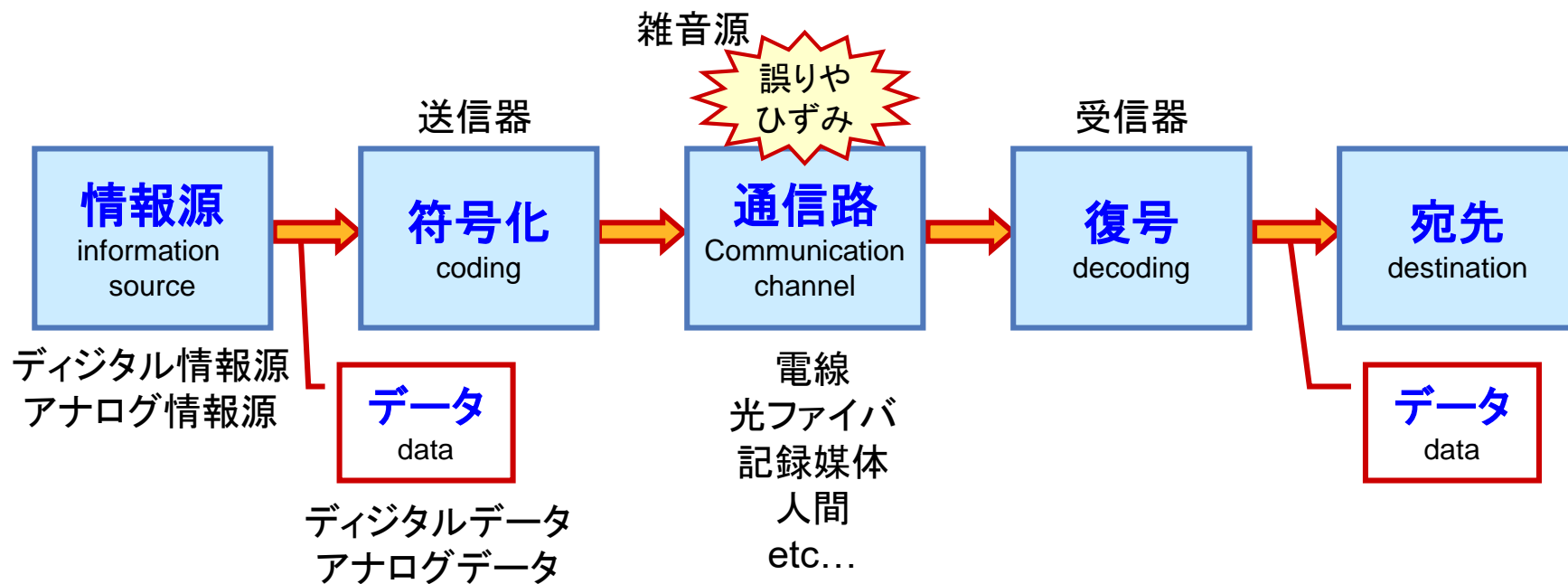


データを圧縮してから伝送すれば、格段に速く(または低コストで)送れる。
ではどこまで圧縮しても大丈夫なのか？ 削れない本質的な情報とは何か？



通信システムのモデル

- ある通信モデルでの情報伝達の効率や信頼性について考える
 - 情報源から発生するデータを、通信路を通して宛先へ伝達する
 - ただし、通信路では雑音やひずみが存在する



この講義では、デジタル情報源・通信路を扱う



情報理論の課題

- 通信路の使用効率
 - 信頼性
- どうやって?
 - 限界は?

問1.2) 天気(晴、曇、雨、雪)を2元通信路(0,1系列を伝える)で送りたい。通信路を使用する際には、送られる0,1の記号数に応じて料金がかかるので、できるだけ送る記号数を減らしたい。どうすればいい?

効率

問1.3) 2元通信路では、困ったことに、1記号ごとに 10^{-3} の確率で0と1が反転する。どうにかして、誤りを少なく記号を送りたい。どうすればいい?

信頼性

表1.2

情報源記号	確率	C1	C2
晴	0.055	0 0	1 0
曇	0.012	0 1	1 1 0
雨	0.002	1 0	1 1 1 0
雪	0.931	1 1	0

符号

符号語

符号アルファベット={0,1} :
2元符号 (q個あったらq元符号)



情報理論の課題

- **情報の量を、いかに客観的に数値化するか
(情報量とエントロピー)**
 - これ以上は圧縮できないという限界を知りたい
- **情報源の性質を利用して、符号化の効率をいかに高めるか
(情報源符号化)**
 - どういう方法で圧縮すると、高い圧縮率を低コストで実現できるか
- **通信路の性能を、いかに客観的に数値化するか
(通信路容量)**
 - これ以上の速度では情報を通せないという限界を知りたい
- **雑音による悪影響を避けるために信頼性をいかに高めるか
(通信路符号化)**
 - どういう方法で符号化すると、誤り検出・訂正を低コストで実現できるか



情報理論の分野

Claude Elwood Shannon (クロード・エルウッド・シャノン)

- C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October, 1948.

符号化の限界を主眼とする理論
シャノン理論ともいう

(狭義の)情報理論

(狭義の)符号理論

暗号理論

安全な情報伝達の理論

(代数学に基づく)通信路符号化の具体的構成法
および符号化・復号法を主眼とする理論

(狭義の)暗号理論は本講義では取り扱わない



今日のまとめ

■ 情報伝達のモデル化

記号列の統計的性質による情報伝達モデル

情報源 → 情報源符号化 → 通信路符号化 → 通信路 →
→ 通信路復号 → 情報源復号 → あて先

■ 情報理論が取り組む4つの課題

【課題1】できるだけよい**情報源符号化法**(復号法)を見出すこと

【課題2】情報源符号化の限界を知ること

【課題3】できるだけよい**通信路符号化法**(復号法)を見出すこと

【課題4】通信路符号化の限界を知ること

■ 情報理論の分野について

次回テーマ: 「情報量とエントロピー」