HERIOT WATT UNIVERSITY

Introduction to Information Theory

Dr. Yoann Altmann

B39AX – Fall 2023
Heriot-Watt University

LEADERS IN IDEAS AND SOLUTIONS

Plan

- Motivations and introduction
- Source model (DMS)
- Information content
- Entropy
- Information rate

- Based on the book:
  “Information Theory,. **Inference**, and **Learning Algorithms”**. David J.C. MacKay

Motivations

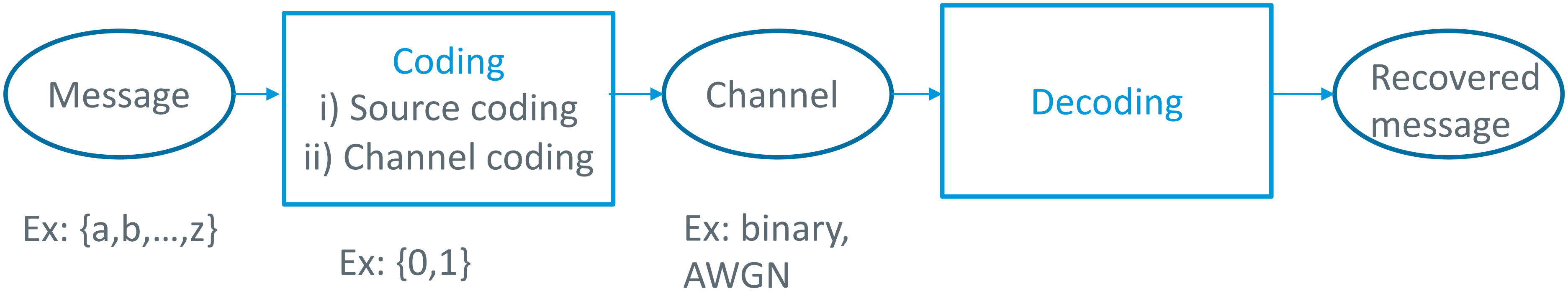
- Noisy communication channels
  - Noise: limits impossible to be overcome?
- Claude Shannon
  - Possibility to transmit information with prob. of error as small as desired provided a suitable code is used
- Important concepts and tools:
  - Quantification of the information?
  - Random noise/errors: probability/statistics

Claude Shannon

- (April 30, 1916 – February 24, 2001) was an American mathematician and electronic engineer known as "the father of information theory"

- Shannon is best known for his 1948 paper "*A Mathematical Theory of Communication*" in which he created the field of Information Theory, but he had many other important contributions in diverse areas.

Typical communication system



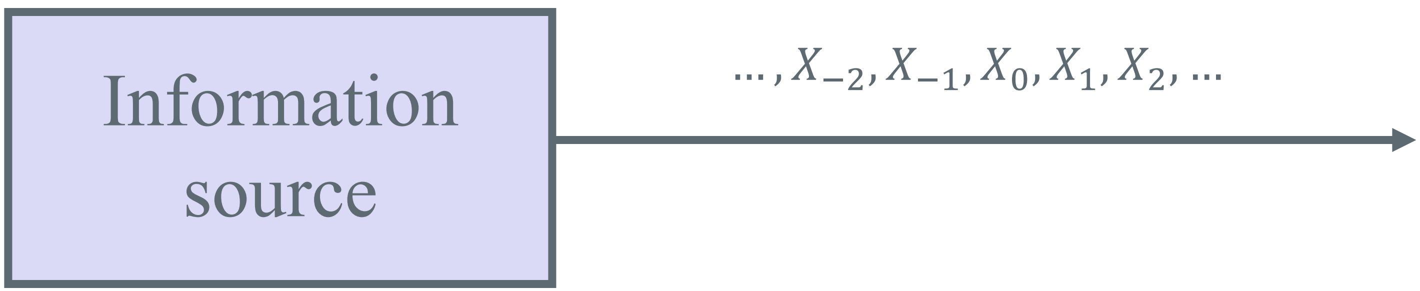
- Coding:
  - Represent the message as efficiently as possible
  - Make the make robust to channel noise
- Joint source/channel coding (~1990) (not covered here)
- 2 types of sources
  - Discrete/continuous

Source/channel coding

- Different objectives
- Source coding
  - Reduce redundancy
  - Shorten the message length
  - Efficient used of bandwidth
- Channel coding
  - Add redundancy
  - Reduce decoding errors
  - Depends on the transmission environment
- Shannon showed that this can be handled separately

Example of source model

- Discrete memoryless source (DMS)
  - A mathematical model for information source: a discrete-time random process $\{X_i\}_{i=-\infty}^{\infty}$



  - The simplest model for the information source is the *discrete memoryless source* (DMS). A DMS is a discrete time, discrete-amplitude random process in which all X_i are generated independently and with the same distribution.

Discrete memoryless source

- X takes values from the alphabet $\mathcal{A} = \{a_1, a_2, \dots, a_N\}$ with probabilities $p_i = \mathbb{p}(X = a_i)$

- A full description of the DMS is given by the set $\mathcal{A}$ (alphabet), and the probabilities $\{p_i\}_{i=1}^{N}$

Examples

- $\mathcal{A} = \{0,1\}$ (binary source)
- $\mathbb{p}(X = 0) = p$, $\mathbb{p}(X = 1) = 1 - p$
- Transmitted data: 001011000100 ...

- $\mathcal{A} = \{00,01,10\}$
- $\mathbb{p}(X = 00) = 0.3$, $\mathbb{p}(X = 01) = 0.5$, $\mathbb{p}(X = 10) = 0.2$
- Transmitted data: $0010\cancel{11}000100$ ...

Impossible: not in alphabet

Memoryless source

- Property: all the symbols are mutually independent

$$\mathbb{p}(X_1, X_2, \dots, X_N) = \mathbb{p}(X_1)\mathbb{p}(X_2) \dots \mathbb{p}(X_N)$$

- Other source model: Markovian source
  - Each symbol depends of the previous one

$$\mathbb{p}(X_1, X_2, \dots, X_N) = \mathbb{p}(X_1)\mathbb{p}(X_2|X_1) \dots \mathbb{p}(X_N|X_{N-1})$$

  - Not covered in this course

Different source models

- Zero-order approximation (symbols independent and equiprobable)

  XFOML RXKHRJFFJUJ ZLPWCFWKCYJ

- First-order approximation (symbols independent but with frequencies of English text)

  OCRO HLI RGWR NMIELWIS EU LL

- Second-order approximation (digram structure as in English)

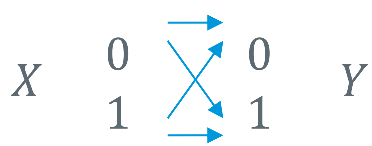
  ON IE ANTSOUTINYS ARE T INCTORE ST BE

- Third-order approximation (trigram structure as in English)

  IN NOT IST LAT WHEY CRATICT FROURE BIRS GROCID

Back to the binary source

- $\mathcal{A} = \{0,1\}$ (binary source)
- $\mathbb{p}(X = 0) = p$, $\mathbb{p}(X = 1) = 1 - p$
- Transmitted data: 001011000100 ...
- Binary symmetric channel



$$\mathbb{p}(Y = 0|X = 0) = 1 - f, \mathbb{p}(Y = 1|X = 1) = 1 - f$$
$$\mathbb{p}(Y = 1|X = 0) = f, \mathbb{p}(Y = 0|X = 1) = f$$

Back to the binary source

- f: Probability of error of a symbol (does not depend on the actual symbol)

- Problem: for $f = 0.1$, 10% of the symbols are corrupted!

$$\mathbb{p}(Y \neq X) = \mathbb{p}(Y = 1|X = 0)\mathbb{p}(\mathrm{X} = 0) + \mathbb{p}(Y = 0|X = 1)\mathbb{p}(\mathrm{X} = 1) = fp + f(1 - p) = f$$

- How can this be improved?

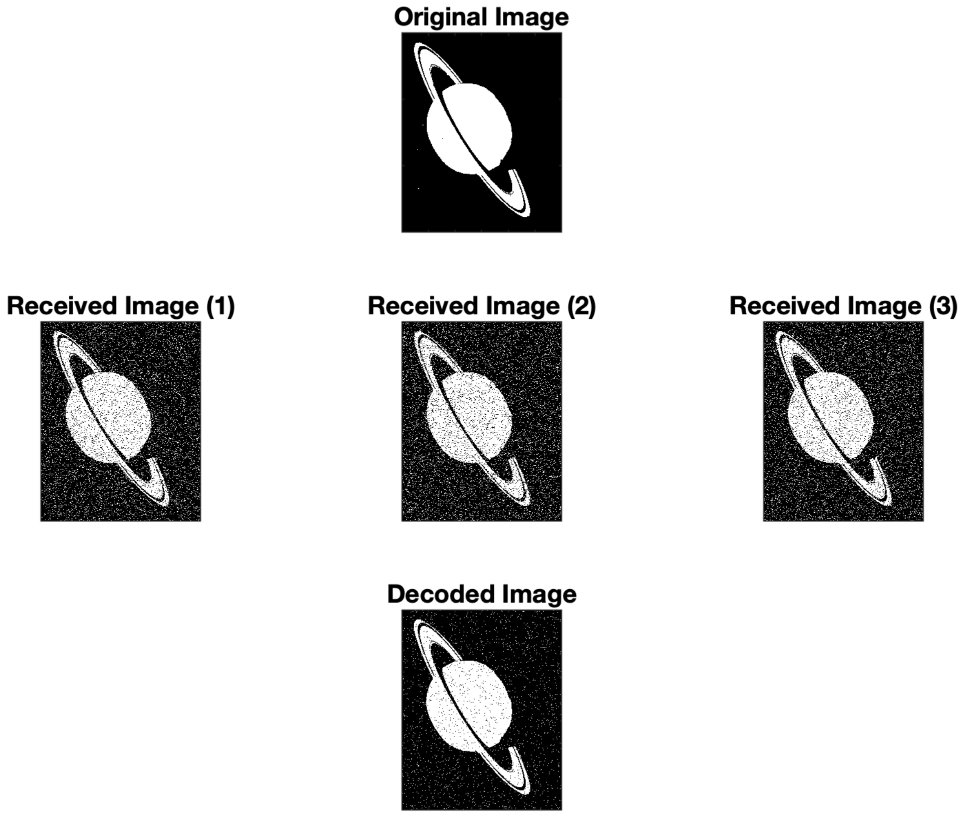
Example: repetition code

- “Naïve” approach: repeated code
- If we send the data twice, we can detect some errors, but not correct them all

$0 \rightarrow 00$, if 01 is received should we decide 0 or 1?

- R_N repetition: demo1.m

Repetition code



Repetition code

- $\mathcal{A} = \{0,1\}$ (binary source)
- $\mathbb{p}(X = 0) = 0.5$, $\mathbb{p}(X = 1) = 0.5$ $(p = 0.5)$
- $f = 0.1$
- Using the R_3 coding/decoding strategy, by how much can we improve the error?
- Demo1.m (+ tutorial)