Mansoura University
Faculty of Computers and Information
Department of Information System

[IS311T] Information Theory

Grade: 3^{rd} grade (IS)

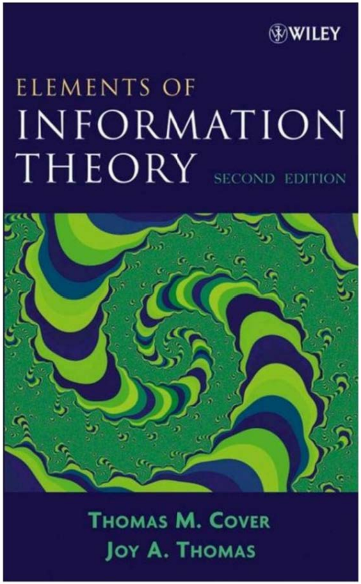
Lecture 02: Introduction

Dr. Islam Reda

Outline

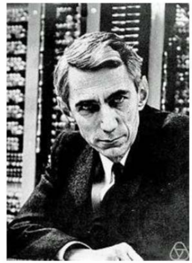
- Textbook
- What is information theory?
- Course syllabus
- What is information?
- The amount of information of an information source
- Entropy
- Examples

Textbook


WILEY
ELEMENTS OF
INFORMATION
THEORY
SECOND EDITION
THOMAS M. COVER
JOY A. THOMAS

What is information theory?

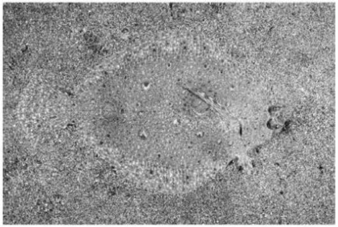
- The central problem in information theory is efficient and reliable transmission of data from a source to a destination.
- Who started it?
  - Claude Shannon (1916 – 2000). The father of information theory.
  - Article in 1948: The mathematical theory of communication.
- Answers to two main question:
  - How much can data be compressed? (The best possible compression that can be achieved)
  - How fast can data be reliably transmitted over a noisy channel? (For a noisy channel: The best possible error correction that can be achieved)

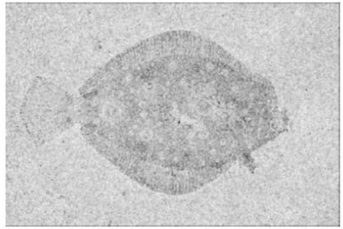


What is information theory?

- Two basic “point-to-point” communication theorems .
  - Source coding theorem: The maximum rate at which data can be compressed losslessly is the entropy rate of the source.
  - Channel coding theorem: The maximum rate at which data can be reliably transmitted is the channel capacity of the channel.
- Applications:
  - Data compression
  - Error correcting coding
  - Transmission and modulation
  - Image processing: texture
  - Information security…

Applications of information theory





17	24	250	8	15
23	5	237	14	16
4	6	244	20	22
10	12	253	251	3
11	18	240	239	9

entropy →

	3.17			

Syllabus

Topics	Required Reading
Introduction	Chapter 1
Entropy and mutual information	Chapter 2.1 – 2.4
Chain rule for entropy and Jensen's inequality	Chapter 2.5 –2.6
Data processing inequality and Fano's inequality	Chapter 2.8, 2.10
Source coding, Kraft inequality, and optimal code length	Chapter 5.1 –5.4
Huffman codes	Chapter 5.6 –5.8
Shannon-Fano-Elias and arithmetic coding	Chapter 5.9 –5.10, 13.3
Overview of channel capacity	Chapter 7.1 –7.4
Channel coding theorem	Chapter 7.5, 7.7
Hamming codes	Chapter 7.11

What is information?

- A sequence of symbols that can be interpreted as a message.
- Information resolves uncertainty. Information is what you get when your uncertainty about something is reduced.
- **Fair coin example**:

  When you flip a fair coin, you are uncertain of whether it will land on heads or tails. You have uncertainty.

  When it lands, your uncertainty is gone.
- **Loss of uncertainty $\equiv$ Gain in information.**

What is information?

- **Unfair or biased coin example**:

  A coin that always lands on heads and never lands on tails.

  When you flip this coin, you have no uncertainty.

  You know in advance that it will land on heads.

  When it lands on heads, you aren't getting any information because your uncertainty hasn't been reduced.

- The more uncertain an event is, the more information is required to resolve uncertainty of that event.

Amount of information

- By calculating the **amount of uncertainty** we have about an information source, we are also calculating the **amount of information** that we will receive when we lose that uncertainty.
- Let X be an information source with M possible outcomes. The amount of information $I(X)$ is given by:

$$\boldsymbol{I(X) = \log(M)}$$

The information received from an information source equals the logarithm of the number of possible outcomes.

- The unit used to measure $I(X)$ depends on the base of the logarithm.

Amount of information

- The base of logarithm used in this course is base 2
- The amount of information is measured in bits.
- Fair coin example:

$$I(X) = \log_2 2 = 1\ bit$$

- For a fair coin with two possible outcomes, we have one bit of uncertainty. When it lands we lose that uncertainty and gain one bit of information.
- Unfair coin Example:

$$I(X) = \log_2 1 = 0\ bit$$

- For an unfair coin that always lands on heads, it has only one possible outcome.

Amount of information

- A dice example:

  A dice has six possible outcomes.

$$I(X) = \log_2 6 = 2.585\ bits$$

- Since the last example has more outcomes, the amount of uncertainty increases. Therefore, the amount of information is bigger than that of the coin examples.

$$\log_2 a = \frac{\log a}{\log 2} \approx 3.32 \times \log a$$

- Important note: For simplicity, we will use (log) instead of ($\log_2$), while we actually mean ($\log_2$).
- In all the previous examples, the outcomes are equally likely.

Amount of information

- What if the outcomes are not equally likely?
- Let $symbols = \{a_1, a_2, \dots, a_M\}$ be the possible outcomes of an information source, the probabilities associated with these symbols be $\{p_1, p_2, \dots, p_M\}$. The amount of information that we receive when we get a specific symbol a_i is:

$$I(a_i) = \log(\frac{1}{p_i})$$

- Examples:

1. A fair coin with a set of two probabilities $\{0.5, 0.5\}$ where the first probability is the probability of getting head and the second one is the probability of getting tail.

$$I(a_1) = \log\left(\frac{1}{0.5}\right) = 1\ bit$$

Amount of information

- Examples:
  2. An unfair coin that always comes up as heads, the probability is {1}

$$I(a_1) = \log\left(\frac{1}{1}\right) = 0\ bit$$

  3. A biased coin with probabilities of {0.99, 0.01}

$$I(a_1) = \log\left(\frac{1}{0.99}\right) = 0.014\ bit$$

$$I(a_2) = \log\left(\frac{1}{0.01}\right) = 6.644\ bit$$

- The outcomes that are less common give us more information and vice versa.

Entropy

- What if we want to measure the average uncertainty for an information source?
- The name for this measurement is entropy.
- The entropy of an information source (X) is called $H(X)$

$$H(X) = \sum_{i=1}^{M} p_i \log\left(\frac{1}{p_i}\right) = -\sum_{i=1}^{M} p_i \log(p_i)$$

- Entropy is the average number of bits required to represent an information source.
- Examples

1. Fair coin with probabilities of {0.5, 0.5}

$$H(X) = 0.5 \times \log(\frac{1}{0.5}) + 0.5 \times \log(\frac{1}{0.5}) = 1\ bit.$$

Entropy

- Examples

2. Biased coin with probabilities of {0.75, 0.25}

$$H(X) = 0.75 \times \log(\frac{1}{0.75}) + 0.25 \times \log(\frac{1}{0.25}) = 0.811\ bits.$$

For every time we flip this biased coin, we will get an average of 0.811 bits of information.

3. Unbalanced dice with probabilities of $\{0.1, 0.1, 0.1, 0.5, 0.1, 0.1\}$

$$H(X) = 5 \times (0.1 \times \log(\frac{1}{0.1})) + 0.5 \times \log\left(\frac{1}{0.5}\right) = 2.161\ bit$$

THANK YOU