2022年資訊理論與編碼技巧：授課方式, 內容與注意事項

Professor Ja-Ling Wu

Depart. of Computer Science and Information Engineering

Graduate Institute of Networking and Multimedia

National Taiwan University

1

為因應新冠肺炎之影響，本課程— 資訊理論與編碼技巧，本學期授課方式與內容將採以下方式進行。

1. 授課方式採線上(直播)及實體 (預錄)並行方式進行

   a. 線上直播部分：利用校方提供之 U-meeting (or Google Meet)進行

      a.1:講員按課綱進行每單元課程之講授或重點說明
      a.2:回覆同學對前次單元所提出的提問

   b. 預錄部分：本課程講授內容將全數以MPEG-4 格式錄製完畢，並儲存於課程網頁中。同學可以任何MPEG-4 compliant decoder 播放詳細課程內容。

2

2. 每週同學在教室上課時間（包括直播之重點說明與問題回覆）原則上以兩小時半為原則，縮短同學群聚之時間而講員（或助教）則至少有一人全程都在教室 （或資訊館214室） 。

3. 經過實地測試，利用MPEG-4 decoder 之快速解碼功能，以1.3x 至1.5x 倍速度播放預錄內容，可以有較佳的聽覺體驗。

4. 任何對課程內容之提問，請於講授該週週五下午五點以前，以電子郵件方式寄給課程助教，以便講員於下週直播時回覆提問。

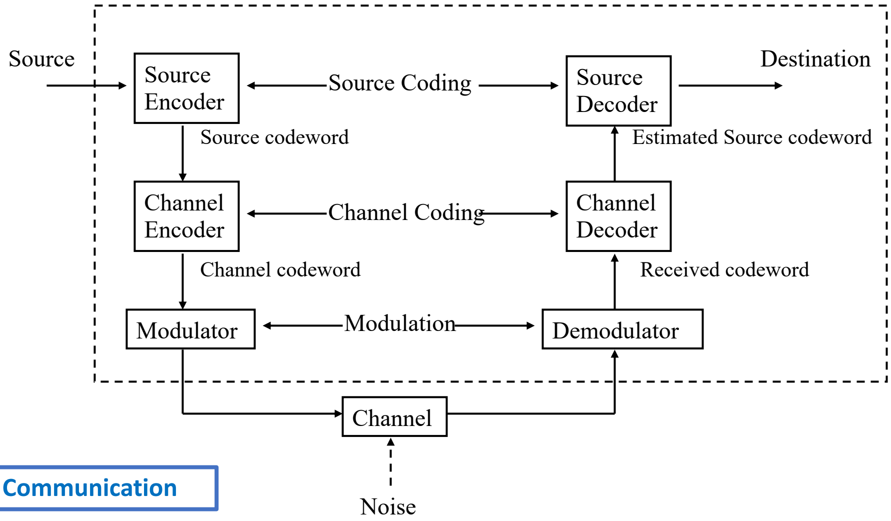
5. 為提升學習效果，建議修課同學提前對預錄之授課內容進行預習。尤其是課程中許多相關數學證明部分，不適宜以直播方式進行，修課同學務必要對預錄之授課內容仔細聆聽。

3

Fundamental Concept of This Course :

4

Digital Communication System



Point –to–Point Communication

Source Coding

Based on characteristics/features of a source, Source Encoder-Decoder pair is designate to reduce the source output to a Minimal Representation. (in terms of what ?)

[**Shannon 1948**]

How to **model** a signal **source**? Random process

How to **measure** the content of a **source**? Entropy

How to **represent** a **source**? Code-design

How to **model** the behavior of a **channel**? Stochastic mapping

Channel capacity

How to **recover/generate** the **best output**? Optimal Filtering

Source Coding (cont.)

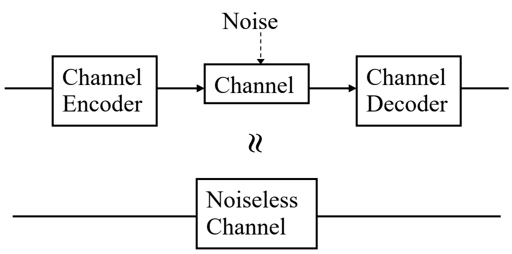
Redundancy Reduction → Data Compression
Data Compaction

Modalities of Sources: (Source's **characteristic** matters!)

- Text
- Image
- Speech/Audio
- Video
- Graphics
- Hybrid

Channel coding

Introducing **redundancy** into the channel encoder and using this redundancy at the decoder to reconstitute the input sequences as accurately as possible, i.e., **Channel Coding** is designate to **minimize the effect/interference of the channel noise**.



Modulation

- Physical channels can require electrical signals, radio signals, or optical signals.
- The **modulator** takes the **channel encoder/source encoder outputs** into account and **transfers the output waveforms** that suit the **physical nature** of the **channel**, and are also chosen to yield either **system simplicity** or **optimal detection performance**.

What is information?

- **What is meant by the "information" contained in an event (a signal source)?**

If we are formally to defined a **quantitative measure** of information contained in an event (or a signal), this measure should have some **intuitive properties** such as:

Characteristics of an Information Measure:

1. Information contained in events ought to be defined in terms of some measure of the uncertainty of the events.
2. Less certain events ought to contain more information than more certain events.
3. The information of unrelated/independent events taken as a single event should equal the sum of the information of the unrelated events.

11

A nature measure of the **uncertainty** of an event α is the **probability of α** denoted $P(\alpha)$.

Once we agree to define the information of an event α in terms of $P(\alpha)$, the properties (2) and (3) will be satisfied if the information in α is defined as

$$\underbrace{I(\alpha)}_{\text{Self-information}} = -\log P(\alpha)$$

＊The **base of the logarithm** depends on **the unit of information** to be used.

12

Information Unit:

$\log_2$: bit
$\log_e$: nat
$\log_{10}$: Hartley

base conversions:

$\log_{10}2 = 0.30103,\quad \log_2 10 = 3.3219$
$\log_{10}e = 0.43429,\quad \log_e 10 = 2.30259$
$\log_e 2 = 0.69315,\quad \log_2 e = 1.44270$

$$\log_a X = \frac{\log_b X}{\log_b a} = (\log_a b)\log_b X$$

Information (Source)

$S_1 \quad S_2 \quad \cdots \quad S_q$: Source alphabet
$P_1 \quad P_2 \quad \cdots \quad P_q$: Probability

Facts:

1) The information content (**surprise**) is somewhat **inversely related to the probability of occurrence**.

1) The information content produced from two different **independent** symbols is the **sum** of the information content generated from each separately.

Since the probability of two **independent** choices are multiplied together to get the probability of the **compound event**, it is natural to define **the amount of information** as

$$I(S_i) = \log\frac{1}{P_i} \qquad (\text{or} - \log P_i)$$

$$I(S_1) + I(S_2) = \log\frac{1}{P_1 P_2} = I(S_1, S_2)$$

As a result, we have

Entropy: **Average information content** over the whole alphabet of symbols

$$H_r(S) = \sum_{i=1}^{q} P_i \log_r\left(\tfrac{1}{P_i}\right) \qquad \begin{Bmatrix} S_1 & S_2 & \cdots & S_q \\ P_1 & P_2 & \cdots & P_q \end{Bmatrix}$$

$$= -\sum_{i=1}^{q} P_i \log_r P_i$$

$$H_r(S) = H_2(S)(\log_r 2)$$

* Consider the entropy of a Source can have no meaning unless a **model of the Source** is included. For a sequence of numbers and if we cannot recognize that they are **pseudo-random** numbers, then we would probably compute the entropy based on the **frequency of occurrence** of the individual numbers.

* The **entropy function** involves only the **distribution of the probabilities** — it is a function of a **probability distribution** P_i and does not involve the S_i

Ex: Weather of Taipei

X = {Rain, fine, cloudy, snow} = {R, f, c, s}
P(R) = ¼, P(F) = ½, P(C) = ¼, P(S) = 0
$H_2(X)$ = 1.5 bits/symbol

If ¼ for each P(i) ⇒ $H_2(X)$ = 2 bits/symbol. (>1.5)
(equal probability event)

H(X) = 0 for a certain (or impossible event)
$P(a_i)=0$ or $P(a_i)=1$

The **logarithmic measure** is more convenient for various reasons:

1. It is practically more useful.
   Parameters of engineering importance such as **time**, **bandwidth**, **number of relays**, etc., tend to vary linearly with the **logarithm** of the **number of possibilities**.
   For example, adding one **relay** to a group doubles the number of **possible states** of the relays. It adds 1 to the base 2 logarithm of this number.

2. It is nearer to the **perception feeling** of a human body

   Light Intensity — eye
   Sound volume — ear

3. It is mathematically more suitable

   $\log_2$ — bits
   $\log_{10}$ — decimal digits
   $\log_e$ — natural unit

   Change from the base a to base b merely requires multiplication by $\log_b a$

Course Outlines

1. Lecture 1 (33mins)

   What is Information Theory? and Why you should take this course?

2. Lecture 2 (2 hrs and 23 mins)

   Entropy, Relative Entropy, and Mutual Information

3. Lecture 3 (1 hr and 46 mins)

   Asymptotic Equipartition Property
   Entropy Rate

4. Lecture 4 (24 mins +56 mins +42 mins)

   Storage can also be modeled as a communication process

   4.1 Channel and Channel Capacity
   4.2 Binary Symmetric Channel
   4.3 Markov Process and Source with Memory

21

5. Lecture 5 (1hr and 20 mins)

   Introduction to Data Compression (increase communication efficiency and reduce storage requirement)

6. Lecture 6 (1 hr and 33 mins + 47 mins)

   6.1 Huffman Codes
   6.2 Research Topics on Huffman Codes (→ Lecture 9)

7. Lecture 7 (1 hr and 31 mins +22 mins +8 mins)

   7.1 Arithmetic Coding
   7.2 Implementation of Arithmetic Coding
   7.3 Secure Arithmetic Coding
   new: Arithmetic Coding in Joint Compression, Encryption and Authentication

22

8. Lecture 8 (1 hr and 24 mins +52 mins +16 mins)

8.1 Lempel-Ziv Coding
8.2 Dictionary Codes and LZ Coding
8.3 Adaptive Dictionary Compression Algorithm
new : How to Speed up Machine Learning Using LZW Coding?
Suggested References for Advanced Study in LZ and Tunstall Codes

9. Lecture 9 (1 hr and 54 mins +1 hr and 35 mins +50 mins)

9.1 Image Data Compression (1)
9.2 Image Data Compression (2)
9.3 Predictive Coding

23

10. Lecture 10 (42 mins +39 mins +30 mins +55 mins)

10.1 Transform Coding Techniques
10.2 Discrete Cosine Transform
10.3 Modified Discrete Cosine Transform
10.4 Fast Algorithm for DCT

11. Lecture 11 (1 hr 21 mis + 1 hr 8 mins +54 mins)

11.1 Overview of Video Coding Algorithms
11.2 Motion Estimation for Video Coding Standards
11.3 Video Compression (Cited from Yao Wang)

12. Lecture 12 (60 mins)

12.1 Rate Distortion Function and Optimal Bit Allocation
new: Data Compression, data Security and Machine Learning

24

Optional Topics :

13. Lecture 13 (1 hr 46 mins + 1 hr 19 mins +50 mins)

13.1 On Compression Encrypted Data (1)
13.2 On Compression Encrypted Data (2)
13.3 On Compression Encrypted Images

14. Lecture 14 (1 hr and 30 mins)

Introduction to Distributed Video Coding
new: Perceptual DVC – Highly Parallel Codes

15 Lecture 15 (1 hr 6 mins + 1 hr 30 mins)

15.1 Information Theory in Data Visualization (1)
15.2 Information Theory in Data Visualization (2)

16. Lecture

Information Theory in Machine learning and Steganography

25

Text Book and References:

- 1. **Elements of Information Theory**, Thomas M. Cover & Joy A. Thomas, Wiley 2nd Edit. **2006**
- 2. Introduction to Data Compression, Khalid Sayood, 1996.
- **3. JPEG/ MPEG-1 coding standard**
- 4. Introduction to Information Theory and Data Compression, Greg A. Harris & Peter D. Johnson, CRC Press, 1998

26

- 5. The **Minimum Description Length** Principle, Peter D. Grunward, the MIT Press, 2007.

- 6. IEEE Trans. on
  Information Theory, **Circuits and Systems for Video Technology**,
  Signal Processing, Image Processing, Multimedia,
  Information Forensics and security, Communications,
  and Computers

- 7. Open Journal : **Entropy**

27

Course Grading

28

Grading: (2020, 2021, 2022, 2023)

Home works (20%)

Mid-term :

1. A brief **Survey** report about any one of the **Information Theory-related Applications** may or may not be covered in the lectures. (**10** %) → **Max 4 pages A4**
2. A brief **Survey** report of the state-of-the-art research status associated with the application chosen above. **(20%)** → **Min 6 pages A4**

Final: (**50%**)

- A group (with Maximum **3 members**) tries to **realize or investigate** an **IT-related final project proposed** by the group members.
- The grading depends on the satisfaction of **the group's ≥30 mins. recorded pptx-based Oral presentation** associate with the **Final hand-in report**. (**Final Report → Min 12 pages A4)**

29

Grading: Basic Requirements : (before 2019)

- 1. Home works (20%)
- 2. A Brief Survey of your interested IT-related Research topic (Minimum: A4 6 pages 15%)
- 3. Programming Assignments (25%)
  - Variable Length Codes
  - Real-Time JPEG Compressor/Decompressor
  - An Optimal JPEG Quantization Table wrt. a given Image

Grade: from C+ to B+ (Survey Quality dependent)

- 4. Final Project: (40%)

30

Examples of Selected Topics for the Final Project:

- 1. **Realization of Video Codecs:**

a Real time Standard Video Decoder

----

b Near real time Standard Video Encoder

----

c Implementation of advanced Standard Multimedia Codecs (H.264/265)

----

d DNN/CNN realization for Standard Video and Image Codecs

31

Examples of Other advanced Multimedia Codecs :

1. Compression Algorithms for **3D** Videos
2. **Parallelized** Video Codecs – **GPU**-based
3. **Scalable** Video Codecs – **Multi-resolution**
4. **Distributed** Video Codecs – **Cloud**-based
5. Advanced **Audio** Codecs
6. Compression Algorithms for **AR/VR** (Metaverse)
7. Compression Algorithms in the **Encrypted Domain**
8. **DNN-based Compression** Approaches for Various Modalities

32

2. Applications of Information Theory in Specific Research Field (1) :

- **a** Survey of **Non-Shannon Entropies** and their Applications (e.g. **Permutation Entropy**)
- **b** Information **Geometric and Topology**
- **c** Information in **Complexity Theory**
- **d** Information theory (Statistics) based **Machine Learning**
- **e** Information Theory in **Security and Privacy Protection**
- **f** Information Theory in **Data Visualization**

2. Applications of Information Theory in Specific Research Field (2):

- **g** Information Theory in **Non-Engineering** Applications
- **h** Information Theory in **Computer Vision and Image Processing**
- **i** Information Theoretic **Analysis of DL (Explainable AI)**
- **j** Information Theory in **Finance Analysis**
- **k** Others --- **propose and confirmed**