Lecture 1

What is Information

Peng-Hua Wang

Graduate Institute of Communication Engineering

National Taipei University

Lecture Information

Reference: Stefan M. Moser and Po-Ning Chen, *A Student's Guide to Coding and Information Theory*, Chapter 5.

1. Motivation

What is information

We begin in asking the question:

What is information ?

Example 1

- "Does the number of typhoons exceed 10 this year?"
  - This question has only two possible answers: "yes" or "no".
- "One people in New Taipei City will win the lottery. Who win?"
  - Here, the question has about 3,897,888 possible answers. (Feb. 2011)
- The second answer provides you a much bigger amount of information than the first one.

First thing we learn

- For a question, the number of possible answers r should be linked to “information”.

Example 2(from our reference)

- You observe a gambler throwing a fair dice. There are 6 possible outcomes $\{1, 2, 3, 4, 5, 6\}$. You note the outcome and then tell it to a friend. By doing so you give your friend a certain amount of information.
- Next you observe the gambler throwing the dice *three times*. Again, you note the three outcomes and tell them to your friend.
- Obviously, the amount of information that you give to your friend this time is three times as much as the first time.

Second thing we learn

- "Information" should be additive in some sense

Example 2

- In the first case, we have $r = 6$ possible answers.
- In the second case, we have $r = 6^3 = 216$ possible answers.
- We have 36 times more possible outcomes in the second case than in the first. But we would like to have only a 3 times larger amount of information!
- How do we match our "feeling" ?
- Solution: Obtain the exponent of the possible answers. That is, take logarithm.
- The researcher Ralph Hartley has made exactly these observations in 1928 in Bell Labs.

Hartley's information

- Definition. We define the following measure of information:

$$\tilde{I}(U) \triangleq \log_b r$$

  where r is the number of all possible outcomes of a random message U.

- We have desired additivity of information

$$\tilde{I}(U_1, U_2, \ldots, U_n) = \log_b r^n = n \log_b r = n\tilde{I}(U)$$

- The base b of logarithm decides about the unit of information. Not really important in our measure.

Base of logarithm

- If $b = 2$, the unit is *bit*.
- If $b = e = 2.71828...$, that is, we take the natural logarithm, the unit is *nat*.
- If $b = 10$, the unit is *Hartley*.
  - In honor of Hartley. It seems that no one use this unit for measuring information ...

Example 3

- There are 103710 people living in San-Shia. How long must a telephone number be if we want to be able to connect to every person?
- $\log_2 103710 = 16.66$. Therefore, we need 17 bits for our phone number.
- $\log_{10} 103710 = 5.01$. We need 6 Hartleys for our phone number. (In fact, we do not need evaluate the logarithm. Just count the number of the decimal digits since we use radix-10 system.)

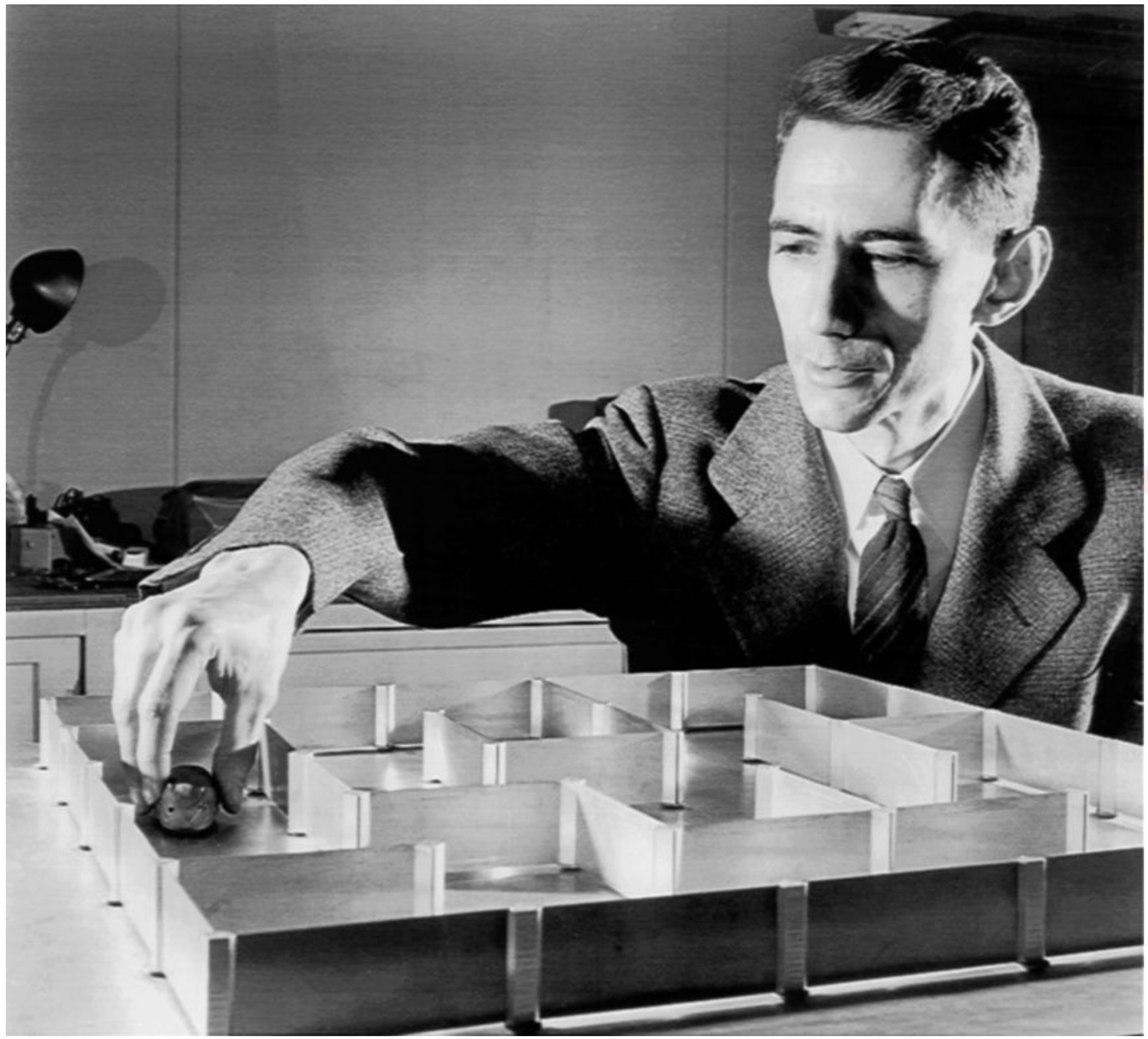
Example 4. Flaw of Hartley information

- Example. Suppose that we have two boxes. Box 1 contains 8 black balls and 8 white balls. Box 2 contains 1 black balls and 15 white balls. We draw one ball at random and let U be the color of the ball.
- By Hartley's information, $\tilde{I}_1 = \log_2 2 = 1 bits$. $\tilde{I}_2 = \log_2 2 = 1 bits$.
- Obviously, U_2 contains less information since white ball is much more likely.

Third thing we learn

- A proper measure of information needs to take into account the probabilities of the various possible events!
- This has been observed for the first time by Claude Elwood Shannon, Father of Information Theory, in 1948 in his landmark paper: “A Mathematical Theory of Communication”.

C. E. Shannon, 1916-2001



Claude Shannon and his clever electromechanical mouse

Example 4

- For Box 2, there is one chance out of 16 possibilities that we draw a black ball. In this case, the Hartley information is

$$\log_2 16 = 4$$

- There is 15 chance out of 16 possibilities that we draw a white ball. That is, 1chance out of 16/15 possibilities. In this case, the Hartley information is

$$\log_2 16/15 = 0.093$$

Example 4

- Now we have two different measure of information depending on the color. How shall we combine them to one value that represents the information? Averaging according to their probabilities :

$$\frac{1}{16}\log_2 16 + \frac{15}{16}\log_2 16/15$$

Shannon's measure of information

$$\sum_{i=1}^{r} p_i \log_b \frac{1}{p_i} = -\sum_{i=1}^{r} p_i \log_b p_i$$

where p_i is the probability of the ith outcome.

- This is the average Hartley information.

2. Uncertainty and Entropy

Entropy

Definition. The uncertainty or entropy of a random message U that takes $r$different values with probability p_i, $i = 1, ..., r$, is defined as

$$H(U) \triangleq = -\sum_{i=1}^{r} p_i \log_b p_i$$

- We now formally define the Shannon’d measure of “self-information of a source”. (i.e. we also have “mutual information”)
- Shannon called his measure entropy due to its relationship with a corresponding concept in physics.
- “uncertainty” would be a far more precise description.

Entropy

- Be careful. Do not to confuse uncertainty with information!
- In Section 1 we have talked a lot about “information”. This is for motivation purposes.
- In fact, what we actually meant is “self-information” or more precisely,“uncertainty”.