

Information Theory

Mohamed Hamada

Software Engineering Lab
The University of Aizu

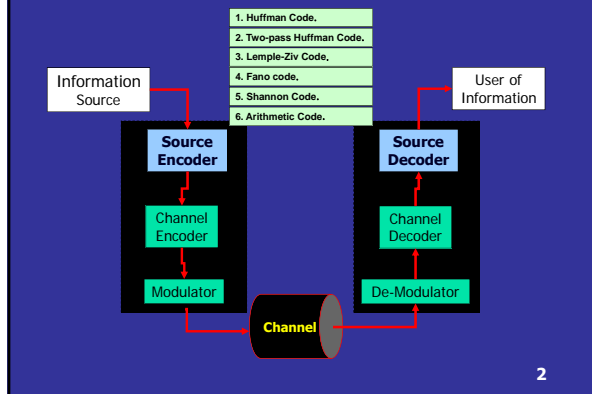
Email: hamada@u-aizu.ac.jp
URL: <http://www.u-aizu.ac.jp/~hamada>

Today's Topics

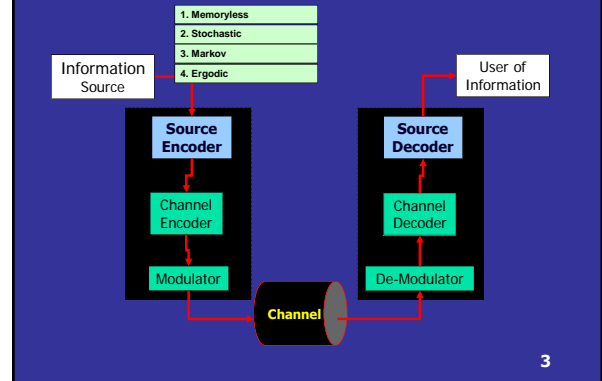
- Communication Channel
- Noiseless binary channel
- Binary Symmetric Channel (BSC)
- Symmetric Channel
- Mutual Information
- Channel Capacity

1

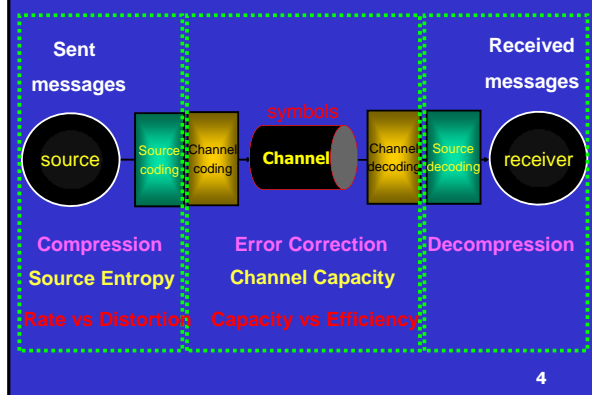
Digital Communication Systems



Digital Communication Systems



INFORMATION TRANSFER ACROSS CHANNELS



Communication Channel

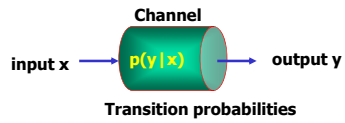
A (discrete) channel is a system consisting of an input alphabet X and output alphabet Y and a probability transition matrix $p(y | x)$ that expresses the probability of observing the output symbol y given that we send the symbol x

Examples of channels:

CDs, CD – ROMs, DVDs, phones,
Ethernet, Video cassettes etc.

5

Communication Channel



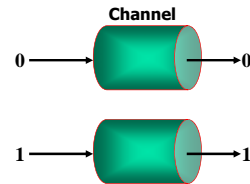
Memoryless:

- output only on input
- input and output alphabet finite

6

Noiseless binary channel

Noiseless binary channel



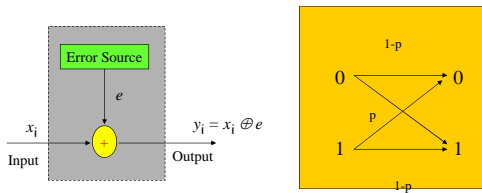
Transition Matrix

$$p(y | x) = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1 & 0 \\ 1 & 0 & 1 \end{array}$$

7

Binary Symmetric Channel (BSC)

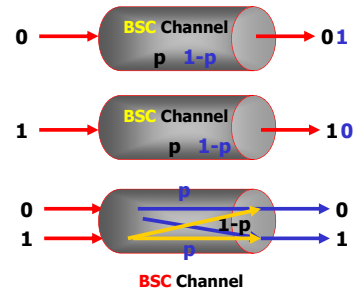
(Noisy channel)



8

Binary Symmetric Channel (BSC)

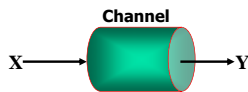
(Noisy channel)



9

Symmetric Channel

(Noisy channel)



In the transmission matrix of this channel, all the rows are permutations of each other and so the columns.

Example:

Transition Matrix

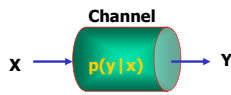
$$p(y | x) = \begin{array}{c|ccc} & y_1 & y_2 & y_3 \\ \hline x_1 & 0.3 & 0.2 & 0.5 \\ x_2 & 0.5 & 0.3 & 0.2 \\ x_3 & 0.2 & 0.5 & 0.3 \end{array}$$

10

Mutual Information

11

Mutual Information (MI)



MI is a measure of the amount of information that one random variable contains about another random variable.

It is the reduction in the uncertainty of one random variable due to the knowledge of the other.

For the two random variable X and Y with a joint probability $p(x, y)$ and marginal probabilities $p(x)$ and $p(y)$.

The mutual information $I(X, Y)$ is the relative entropy between the joint distribution and the product distribution $p(x)p(y)$ i.e.

$$I(X, Y) = \sum_x \sum_y p(x, y) \log_2 \left(\frac{p(x, y)}{p(x)p(y)} \right)$$

12

Mutual Information (MI)

Relationship between Entropy and Mutual Information

$$I(X, Y) = H(Y) - H(Y|X)$$

Proof:

$$I(X, Y) = \sum_x \sum_y p(x, y) \log_2 \left(\frac{p(y|x)}{p(y)} \right)$$

$$I(X, Y) = -\sum_x \sum_y p(x, y) \log_2(p(y)) + \sum_x \sum_y p(x, y) \log_2(p(y|x))$$

$$I(X, Y) = -\sum_y p(y) \log_2(p(y)) + \sum_x p(x) \sum_y p(y|x) \log_2(p(y|x))$$

$$= H(Y) - H(Y|X)$$

Note that $I(X, Y) \geq 0$

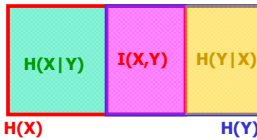
i.e. the **reduction** in the description length of X given Y
or: the amount of information that Y gives about X

13

Mutual Information (MI)

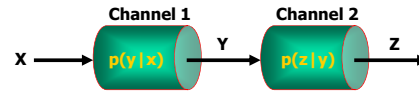
Note That:

- $I(X, Y) = H(Y) - H(Y|X)$
- $I(X, Y) = H(X) - H(X|Y)$
- $I(X, Y) = H(X) + H(Y) - H(X, Y)$
- $I(X, Y) = I(Y, X)$
- $I(X, X) = H(X)$



14

Mutual Information with 2 channels



Let X, Y and Z form a Markov chain: $X \rightarrow Y \rightarrow Z$
and Z is independent from X given Y

$$\text{i.e. } p(x, y, z) = p(x) p(y|x) p(z|y)$$

The amount of information that Y gives about X given Z is:

$$I(X, Y|Z) = H(X|Z) - H(X|YZ)$$

15

Mutual Information with 2 channels

Theory: $I(X, Y) \geq I(X, Z)$

(I.e. Multi-channels may destroy information)

$$\begin{aligned} \text{Proof: } I(X, (Y, Z)) &= H(Y, Z) - H(Y, Z|X) \\ &= H(Y) + H(Z|Y) - H(Y|X) - H(Z|YX) \\ &= I(X, Y) + I(X; Z|Y) \\ I(X, (Y, Z)) &= H(X) - H(X|YZ) \\ &= H(X) - H(X|Z) + H(X|Z) - H(X|YZ) \\ &= I(X, Z) + I(X, Y|Z) \end{aligned}$$

now $I(X, Z|Y) = 0$ (independency)

Thus: $I(X, Y) \geq I(X, Z)$

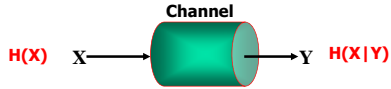
16

Channel Capacity

17

Transmission efficiency

I need on the average $H(X)$ bits/source output to describe the source symbols X . After observing Y , I need $H(X|Y)$ bits/source output.



Reduction in description length is called the transmitted Information.

$$\text{Transmitted } R = I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \text{ from earlier calculations}$$

We can *maximize* R by changing the input probabilities.

The maximum is called **CAPACITY**

18

Channel Capacity

The channel capacity C is the highest rate, in bits per channel use, at which information can be sent with arbitrarily low probability of error

$$C = \max_{p(x)} I(X, Y)$$

Where the maximum is taken over all possible input distributions $p(x)$

Note that :

- During data compression, we remove all redundancy in the data to form the most compressed version possible.
- During data transmission, we add redundancy in a controlled fashion to combat errors in the channel

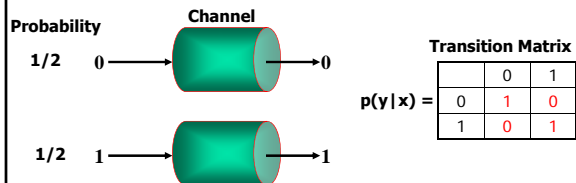
Capacity depends on input probabilities

because the transition probabilities are fixed

19

Channel Capacity

Example 1: Noiseless binary channel



Here the entropy is:

$$H(X) = H(p, 1-p) = H(1/2, 1/2) = -1/2 \log 1/2 - 1/2 \log 1/2 = 1 \text{ bit}$$

$$H(X|Y) = -1 \log 1 - 0 \log 0 - 0 \log 0 - 1 \log 1 = 0 \text{ bit}$$

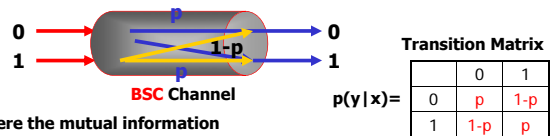
$$I(X, Y) = H(X) - H(X|Y) = 1 - 0 = 1 \text{ bit}$$

20

Channel Capacity

Example 2: (Noisy) Binary symmetric channel

If P is the probability of an error (noise)



Here the mutual information

$$\begin{aligned} I(X, Y) &= H(Y) - H(Y|X) = H(Y) - \sum p(x) H(Y|X=x) \\ &= H(Y) - \sum p(x) H(P) = H(Y) - H(P) \sum p(x) \quad (\text{Note that } \sum p(x) = 1) \\ &= H(Y) - H(P) \quad (\text{Note that } H(Y) \leq 1) \\ &\leq 1 - H(P) \end{aligned}$$

$$I(X, Y) \leq 1 - H(P)$$

Hence the information capacity of the BSC channel is

$$C = 1 - H(P) \text{ bits} \quad (\text{maximum information})$$

21

Channel Capacity

Example 3: (Noisy) Symmetric channel

The capacity of such channel can be found as follows:

let r be a row of the transition matrix

$$\begin{aligned} I(X, Y) &= H(Y) - H(Y|X) = H(Y) - \sum p(x) H(Y|X=x) \\ &= H(Y) - \sum p(x) H(r) = H(Y) - H(r) \sum p(x) = H(Y) - H(r) \\ &\leq \log |Y| - H(r) \end{aligned}$$

$$\text{Hence the capacity } C \text{ of such channel is: } C = \max I(X; Y) = \log |Y| - H(r)$$

For example: Consider a symmetric channel with the following transmission matrix

It is clear that such channel is symmetric since rows (and columns) are permutations.

Hence the capacity C of this channel is:

$$C = \log 3 - H(0.2, 0.3, 0.5) = 0.1 \text{ bits (approx.)}$$

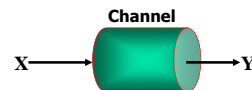
	Y_1	Y_2	Y_3
X_1	0.3	0.2	0.5
X_2	0.5	0.3	0.2
X_3	0.2	0.5	0.3

Transition Matrix

22

Channel Capacity

(Noisy) Symmetric channel



Note that: In general for symmetric channel the capacity C is

$$C = \log |Y| - H(r)$$

Where Y is the output alphabet and r is any row in the transmission matrix.

23