Information Theory and Source Coding

Introduction : Information theory provides a quantitative measure of the information contained in message signal and allows us to determine the capacity of a communication system to transfer this information from source to destination. ln this section we brietly explore some basic ideas involved in information theory and source coding.

Measure of Information

1) Informatiion sources:
An information source is an object that produces an event. the outcome of which is selected at random accorrding to a probability distribution. A practical source in a communication system is a device that produces messages. and it can be either analog or discrete.
In this chapter we deal mainly with the discrete sources, since analog sources can be transformed to discrete sources through the DSC of sampling and quantization techniques, described in Chap. 5.

A discrete information source is a source that has only a finite set of symbols as possible outputs. The set of source symbols is called the **source alphabet**, and the elements of the set are called *symbols* or *letters*.

Memory

Information sources can be classiflied as having memory or being memoryless. A source with memoray is one for which a current symbol depends on the previous symbols. A memoryless source is one for which each symbol produced is independent of the previous symbols.

A discrete memoryless sources- (DMS) can be characterized by the list of the symbols, the probability assignment to these symbols, and the specification ofthe rate of generating these symbols by the source.

Information content of a Discrete Memoryless Source

The amount of infomation contained in an event is closely related to its uncertainty. Messages containing knowledge of high probability ofoecurrence convey relatively little infomation. We note that if an event is certain (that is, the event occurs with probability 1), it conveys zero information.
Thus, a mathematical measure of information should be a function of the probability of the outcome and should satisfy the following axioms:

1. Information should be proportional to the uncertainty of an outcome.
2. Information contained in independent outcomes should add.

Information Content of a Symbol:

Consider a DMS, denoted by X, with alphabet $x, .x2...., x_n$. The information content ofa symbol x_l, denoted by $I(x_i)$, is defineed by

$$I(x_i) = log_b(\frac{1}{P(x_i)}) = -log_b(P(x_i))$$

where $P(x_i)$ is the probability of occurrence of symbol x_i,.

Note that $I(x_i)$ satisfies the following properties;

-
-
-
-

The unit of I(x)) is the bit (binary unit) if b Z 2, Hartley or decit if b = 10, and nat (natural unir) if b Z at It is standard to use b Z 2. Here the unit bit (abbreviated ”b) is a measure ofinlbrrnation content and is not to bc confused with the term hit meaning ”binary digit.” The conversion of these units to other units can be achieved by the following relationships.

Average Information or Entropy

- In a practical communication system, we usually transmit long sequences of symbols from an information source.
- Thus, we are more interested in the average infornation that a source produces than the information content of a single symbol.
- The mean value of $I(x_i)$ over the alphabet of source X with tn different symbols is given by

Entropy

- The quantity $H(X)$ is called the entropy of source X.
- It is a measure of the average information content per random symbol.
- The source entropy H(X) can be considered as the average amount of uncertainty within source X that is resolved by use of the alphabet.
- Note that for if binary source X that generates independent symbols 0 and 1 with equal probability, the source entropy $H(X)$ is
- The source entropy$H(X)$ satisfies the following relation: 0 Z H(X) Z log; in (l0.9) where in is the size (number of symbols) of the alphabet of source X).
- The lower bound corresponds to no uncertainty, which occurs when one symbol has probability $P(>:,)Zl$while P(x,) = 0 for j ye i, so X emits the same symbol x, all the time.
- The upper bound corresponds to the maximum uncertainty which occurs when P(x,) : 1 /1n for all 1. that is, when all symbols are equally likely to be emitted by X.

Information Rate

If the time rate at which source X emits symbols is r (symbols/s), the Information rate R of the source is given by

DISCRETE MEMORYLESS CHANNELS

A. Channel Representation:

- A communication channel is the path or medium through which the symbols flow to the receiver.
- A discrete memoryless channel (DMC) is a statistical model with an input X and an output Y. During each unit ofthe time (signaling interval), the channel accepts an input symbol from X, and iu response it generates an output symbol from Y.
- The channel is ”discrete” when the alphabets of X and Y are both finite.
- It is “memoryless” when the current output depends on only the current input and not on any of the previous inputs.

Discrete memoryless channel

A diagram of a DMC with nt inputs and n outputs is illustrated in Fig. l0-1. The input X consists of input symbols ir, , X2, xm. The a priori probabilities of these source symbols P(.x,) are assumed to be known. The output Y consists of output symbols $\{y_1, y_2, \ldots, y_i\}$
Each possible input-to-output path is indimted along with a conditional probability $P(y_i|x_i)$, where$P(y_i|x_i)$ is the conditional probability of obtaining output y_i given that the input is x_i, and is called a ***channel transition probability***.

Channel Matrix

A channel is completely specified by the complete set of transition probabilities. Accordingly, the channel of Fig. 10-l is ol‘ ten speciiied by the matrix of transition probabilities [P(YlX)l, given by
The matrix [P(YlX)] is called thc channel matrix. Since each input to the channel results in some output, each row of the channel matrix niust snm to unity. that is,

MUTUAL INFORMATION

A. Conditional and Joint Entropies:
Using the input probabilities P(x,), output probabilities $P(y_i)$, transition probabilities P(yJ—¿r,), and joint probabilities P(x,, yy), we can define the following various entropy functions for a channel with m inputs and n outputs:

- H(X) = - X P(xi) log
- H(Y) = -2P(yj)]
- H¡X I Y) = - X X
- H = -2 ZP
- H(X, Y) - -2 Z F

Conditional and Joint Entropy

These entropies can be interpreted as follows: H(X) is the average uncertainty of the channel input, and H(Y) is the average uncertainty of the channel output. The conditional entropy H(X]Y) is a measure of the average uncertainty remaining about the channel input after the channel output has been observed. And H(X] Y) is sometimes called the equivncation of X with respect to Y.

- The conditional entropy H(Y—X) is the average uncertainty of the channel output given that X was transmitted.
- The joint entropy H(X, Y) is the average uncertainty of the communication channel as a whole.

Two useful relationships among the above various entropies are

- $H(X,Y) = H(X|Y) + H(Y)(10,26)$ $H(X,Y) = H(Y|X) + H(X)(10.27)$

B. Mutual Information: The mutual information 1(X; Y) of a channel is deiined by I(X; Y) = H(X) H(X—Y) b/symbol (10.28)

Self Information

Self-information This is defined by the following mathematical formula:$I(A) = logbP(A)$
The self-information of an event measures the amount of ones surprise evoked by the event. The negative logarithm $logbP(A)$ can also be written as

$$log_b \frac{1}{P(A)}$$

Note that log(1) = 0, and that $|log(P(A))|$ increases as P(A) decreases from 1 to 0. This supports our intuition from daily experience. For example, a low-probability event tends to cause more “surprise”.

Code efficiency and Code redundancy

The parameter L represents the average number of bits per source symbol used in the source coding process. The code efficiency is defined as

$$\nu = \frac{L_{min}}{L}$$

where L_{min} is the minimum possiblve value of L. When ν approaches unity, the codes is said to be efficient. The code redundancy γ is defined as $\gamma = 1 - \nu$.

Source Coding Theorem

The source coding theorem states that for zi DMS X with entropy $H(X)$, the average code word length L per symbol is bounded as L 2 H(X) (10.52) and further, L can bc made as close to H(X) as dcsircd for some suitably chosen code. Thus, with$L_{min} \geq H(X)$.
The code efficiency can be rewritten as

$$\nu = \frac{H(X)}{L}$$

Kraft inequality

- Let X be a DMS with alphabet ($x_i = \{1, 2, ..., m\}$). Assume that the length of the assigned binary code word corresponding to x, is n,.
- A necessary and sufficient condition for the existence of an instantaneous binary code is

$$K = \sum_{i=1}^{m} 2^{-n_i} \leq 1$$

which is known as the **Kraft inequality**.

- Note that the Kraft inequality assures us of the existence of an instantaneously decodable code with code word lengths that satisfy the inequality. But it does not show us how to obtain these code words, nor does it say that any code that satisfies the inequality is automatically uniquely decodable

10.5. A high-resolution blackand-white TV picture consists of about 2×10^6 picture elements and 16 different brightness levels. Pictures are repeated at a rate of 32 per second. All picture elements are assumed to be independent, and all levels have equal likelihood of occurrence. Calculate the average rate of information conveyed by this TV picture source.