Information Theory and Coding

Introduction – What is it all about?

References:

C.E.Shannon, "A Mathematical Theory of Communication", *The Bell System Technical Journal,* Vol. 27, pp. 379–423, 623–656, July, October, 1948.

C.E.Shannon "Communication in the presence of Noise", *Proc. Inst of Radio Engineers,* Vol 37 No 1, pp 10 – 21, January 1949.

James L Massey, "Information Theory: The Copernican System of Communications", IEEE Communications Magazine, Vol. 22 No 12, pp 26 – 28, December 1984

Mischa Schwartz, "Information Transmission, Modulation and Noise", McGraw Hill 1990, Chapter 1

The purpose of telecommunications is the efficient and reliable transmission of real data – text, speech, image, taste, smell etc over a real transmission channel – cable, fibre, wireless, satellite.

In achieving this there are two main issues. The first is the source – the information. How do we code the real life usually analogue information, into digital symbols in a way to convey and reproduce precisely without loss the transmitted text, picture etc.?

This entails proper analysis of the source to work out the best reliable but minimum amount of information symbols to reduce the source data rate.

The second is the channel, and how to pass through it using appropriate information symbols the source. In doing this there is the problem of noise in all the multifarious ways that it appears – random, shot, burst- and the problem of one symbol interfering with the previous or next symbol especially as the symbol rate increases.

At the receiver the symbols have to be guessed, based on classical statistical communications theory. The guessing, unfortunately, is heavily influenced by the channel noise, especially at high rates.

To get around this problem, techniques are used to improve the reliability of the guess at the receiver. This entails additional symbols worked into the information symbols producing a code of symbols. The receiver can therefore guess at the level of the symbol, but has also additionally the possibility of guessing, somewhat more reliably, at the level of the code. Coding is therefore the art and science of manipulating the traveling data symbols in a way to decrease as far as possible the possibility of passing erroneous information to the receiver (sink).

Two important methods are used. The first gives the ability to the receiver to detect an error. Automatic repeat request (ARQ) methods are then used to tell the source to retransmit that portion which the receiver has detected an error. The second method involves adding enough redundant symbols to the information symbols, to allow not only the capability of detecting, but also that of correcting wrongly received symbols. This is known as Forward Error Correction, and uses various types of codes.

In the above the source coding has been separated out from the channel coding. However there is also the possibility of combining source and channel coding to reduce symbol rate but keeping the same reliability.

Terminology: Of course the above has to translate into scientific method and engineering jargon. This will include, information, entropy, source coding – Huffman Coding, Arithmetic Coding, Dictionary Coding Methods (LZW and others), Shannon's Theorems, channel capacity, bit rates, symbol rates, probability of bit error, linear algebra – groups, rings, vector spaces, primitive roots -, block codes, convolutional codes, probability of block error etc. etc.

Information

The amount of information in a source is related to its probability of occurrence. A message which a priori has a high probability of occurring conveys very little information when it is received. Based on this, the measure of information of an event X occurring with probability P_X from among a set of possible events, is defined as

$$I_X = \log \frac{1}{P_X} \qquad (1.1)$$

If the probability of occurrence is 1, then no information is conveyed, since the receiver already knows what the transmitter is sending.
The minimum number of events necessary, to require the receiver to make a decision, is two events. Given two events, assumed equiprobable, the easiest symbol representation is the binary pulse, which is called the bit. At the level of source coding the possible events occurring from a given data source are converted to bits. For n events there is a need of $\log_2 n$ bits of information (eg PCM in speech). For equiprobability

$$I_i = \log \frac{1}{P_i} \quad = \quad \log_2 n \quad bits \qquad (1.2)$$

In practice the events are read in, coded and transmitted sequentially. If each event takes time τ , and an interval T is considered, the information transmitted over T seconds is

$$Information = \frac{T}{\tau} \log_2 n \ bits \qquad (1.3)$$

Entropy

The information calculated above is based on equiprobability. In practice, events that need to be transmitted are not like this. Ex: Text has some characters more probable than others; voice has more content at low frequencies.
The information in the message now depends on two factors:

The relative frequency of one event defined as

$$P = \frac{\text{number of times event occurs}}{\text{Total number of events}}$$

The average of the information due to each event, also known as the entropy, H of the source, and given by

$$H = \sum_{i=1}^{n} P_i \log_2 \frac{1}{P_i} \qquad (1.4)$$

In practice, the value of information for H in (1.4) is less or equal to the value of I in (1.2) which is itself the average for the equiprobable event.
(see Shannon's paper, for the graph of H vs p, the probability of one event, given the other probability is (1-p)).

Adding in (1.4) the issue of the time, taken for transmission, the value for H becomes

$$H_T = \frac{T}{\tau} \sum_{i=1}^{n} P_i \log_2 \frac{1}{P_i} \; bits\, in\, T \sec onds \qquad (1.5)$$

The entropy of a source therefore depends on the relative frequency of the events it is made up of. Shannon, in his paper defined the ratio of the entropy of a source to the maximum value it could have, while still restricted to the same symbols (events), as its *relative entropy*. For transmission the events need to be coded, usually into bits, and this is the maximum compression possible when we encode into the same alphabet. One minus the relative entropy is the *redundancy*. The redundancy of ordinary English, not considering statistical structure over greater distances than about eight letters, is roughly 50%. Huffman, Arithmetic and other types of data compression methods, (codes) will be considered to address this compression available in a source, to reduce the bit rate.

Shannon's Coding Theorems

Shannon's first theorem is the source coding theorem. It states that every communication source is characterized by a single number H_T, the source rate, and can be represented as accurately as desired by R binary digits per second if only $R > H$. (Note that H is being used instead of H_T as this is used for both meanings ie bits and bits per second)
Thus, the theorem establishes that the entropy rate is the rate of an optimal lossless data compression code. The limit exists as long as the source is stationary. How to achieve the limit is the realm of data compression methods.

A communication system capable of transmitting this information should have the capacity to handle the average information bits generated.

$$C_{AV} \geq -\frac{1}{\tau}\sum_{i=1}^{n} P_i \log_2 P_i \ bits\ per\ \sec ond \qquad (1.6)$$

On the channel itself, the bits can be transformed to other symbols. This gives rise to the baud (symbol) rate over the channel, and is characterized by

$$C = \frac{1}{\tau}\log_2 m\ bits/s \qquad (1.7)$$

where τ is the time to transmit a symbol over the channel
m is the number of symbols available on the channel

(eg for 8 QPSK, there are 8 symbols, and each symbol requires 3 bits.) So (1.7) must be capable of handling (1.6) with an arbitrary small frequency of errors. 1.7 can also be written as

$$C = 2B\log_2 m\ bits/s \qquad (1.8)$$

where B is the bandwidth of the channel

This is Shannon's first theorem, which assumes a noiseless, lossless, channel. The issue of conditional dependence, as opposed to statist6ical independence, of the set of events will be considered in lab work. The issue of lossy data compression and rate-distortion theory will be considered later.

The second theorem is the channel coding theorem.
In practice communication channels are not noiseless. They all exhibit noise. The quality of signals over practical channels is expressed in terms of the **signal-to-noise ratio** expressed in decibels and given by

$$10 \log_{10}\frac{S}{N}\ dB \qquad 1.9$$

The Shannon-Hartley Law states that for a channel with bandwidth B Hz and a signal-to-noise ratio of S/N (not in dB) the maximum number of bits per second is given by

$$\mathrm{C} = B\log_2(1+\frac{\mathrm{S}}{\mathrm{N}})\ \mathrm{bps} \qquad 1.10$$

C is the theoretical tightest upper bound on the information rate. If we have a channel with capacity C and a source whose rate of generating information R is less than or equal to C , it is possible to encode the source in such a way as to transmit over this channel with the fidelity measured by N . Considering again the 8_QPSK system. The noise N adds a sphere around each point. The number of different messages which must be capable of distinct transmission is of the order of the volume of the region of possible messages divided by the volume of the small spheres. Shannon's theorem states that R can still be of the order of C in (1.10). What is important is not the S/N (so long as R<C), but how the information is coded to achieve the theoretical bound. This implies that information should not be transmitted one bit at a time, but as a block of bits coded into a channel input sequence.

Channel Noise

Channel noise can be random or thermal noise or impulse noise due to lightning, cross-talk etc. Random noise gives rise to random-bit errors. Impulse noise gives rise to burst errors (ie a number of adjacent bits in error).

In practice the minimum signal level required for acceptable bit error rates is important. The energy measured in joules per bit in a signal, denoted by E_b, is given by

$E_b = ST_b$, where S is the signal power in watts and T_b the bit time.

The level of thermal noise in a bandwidth of 1 Hz is defined as

$N_0 = kT$ watts/Hz, where k is Boltzmann's constant and T the temperature in ^{0}K. Hence

$$\frac{E_b}{N_0} = \frac{S/R}{kT}$$

In terms of the channel bandwidth, the noise energy $N = WN_0$, so that

$$\frac{E_b}{N_0} = \frac{S}{N}\frac{W}{R}$$

which finally yields in dB

$$\frac{E_b}{N_0}\,\mathrm{dB} = 10\log_{10}\left(\frac{S}{N}\right) + 10\log_{10} W - 10\log_{10} R$$

The signal power required to achieve a required E_b/N_0 increases with temperature and bit rate R.