Report of Shannon's 1948 Paper

WU, Chenhao

Student ID: 117010285

January 19th, 2019

1   Preview

The development of electronic communication systems grew directly from the creations of telegraph and telephone. Through these approaches electronic technologies were applied on these communication systems in order to transfer information and establish communications between senders and receivers. During the early ages of its development engineers and experimenters encountered many electrical problems while sending signals over wires, and even though the pioneers of electronic communication systems realized the problem as practical engineering problems, before Claude Shannon isolated the meaning of transmitting messages from the communication signals and proposed his ways of generalizing factors affecting communicating channel and analyzing communications in mathematical approaches, there was not such a systematic and reliable theory in electronic communication engineering such that can analysis the communications and measure the information of them.

In Claude Shannon's *A Mathematical Theory of Communication*, Shannon first abstracted communication processes as five parts, information source, transmitter, channel, receiver and destination. Then he classified communication systems as four categories:(1) system with discrete noiseless signal, (2) system with discrete signal affected by noise, (3) system with continues noiseless signal and (4) system with continues signal affected by noise. Shannon proposed sets of theorems in his paper and applied them to each of these systems respectively and illustrate his approach of analyzing electronic communication in real world problems and cases.

2 Problems Discussed in the Paper

2.1 Information Theory Problems in Communication

As Shannon firstly pointed out in the introduction of the paper, the fundamental problem of communications is that of reproducing at one point either exactly or approximately a message selected at another point.

First problem in communication is how one can measure the capacity of such a channel to transmit information. For this problem Shannon based on given examples of teletype and telegraphy, pushed forward his definition of the capacity C of a discrete channel by

$$C = \lim_{T \to \infty} \frac{logN(T)}{T}$$

where $N(T)$ is the number of allowed signals of duration T. Using this criteria the capacity of a channel can therefore be calculated.

Since that if the conditions of the channel are stable for a long period the capacity of that channel is determined, another important problem in communication is of that given a constant capacity how one can maximize the quantities of information that receiver received. By cases Shannon realized that the amount of knowledge conveyed by signal does not directly related to the size of message, and he therefore concentrated on the problems associated with sending process and receiving process.

The first case considers the discrete source of information. Problem of this case is that how can an information source to be described mathematically, and how to calculate information in bits per second produced in a given source. Shannon listed bunch of discrete information sources such as natural languages, PCM transmitter, and sequence of symbols which are generated by an abstractly stochastic process. For the last type of discrete information sources Shannon developed the characteristics

of this type of sources by showing series of approximations to English and realize the approximation as a Markov process. Then he introduce the definition of entropy H to describe a discrete channel by

$$H = -K \sum_{i=1}^{n} p_i \log p_i$$

where K is a positive constant, and for continuous channel by

$$H = - \int_{-\infty}^{\infty} p(x) log p(x) dx$$

Also, the joint entropy of multi-variable can be easily derived by applying probability formulas. Using entropy Shannon treats an information source as choices and uncertainties of a process, and the information and rate of the information produced by such a process can therefore be measured.

The second case involved the impact of noise to the channel during transmission, that Shannon gives the representation of a noisy discrete channel in a form of received signal E to be a function of the transmitted signal S and the second variable, the noise N.

$$E = f(S, N)$$

By pointing out the essence of messages transmitting in the communication process and treating both the information conveyed by message and noise in the channel as random variables, Shannon discussed the possible maximum amount of information that receiver is able to receive from a given channel. For both discrete and continuous channel, the transmission rate of message can be given as

$$R = H(x) - H_y(x)$$

where $H(x)$ is the entropy of the input and $H_y(x)$ is the equivocation, and the channel capacity C is defined as the maximum of R when the input over all possible ensembles are vary.

For the discrete noiseless channel, Shannon has proved that if the source have entropy H (bits per symbol) and a channel have a capacity C (bits per second), then it is possible to encode the output of the source at the average rate $\frac{C}{H} - \epsilon$ and it is not possible to transmit at an average rate greater than $\frac{C}{H}$.

For the discrete channel with noise, Shannon also showed that for a channel with the capacity C and a discrete source with entropy H, if $H \leq C$ there exists a coding system such that the output of the source can be transmitted over the channel with an arbitrarily small frequency of errors, and if $H > C$ it is possible to encode the source so that the equivocation is less than $H - C + \epsilon$ where ϵ is arbitrarily small.

For the continuous channel Shannon also developed that if the capacity of a discrete channel can be given by

$$C = \lim_{T \to \infty} \max_{P(x)} \frac{1}{T} \int \int P(x,y) log \frac{P(x,y)}{P(x)P(y)} dxdy$$

It has been proved in the paper that if the signal and noise are independent and the received signal is the sum of the transmitted signal and the noise then the rate of transmission is

$$R = H(y) - H(n)$$

The maximum capacity in cases include various forms of noise are also discussed in the paper, such as white thermal noise, Gaussian noise and etc.

Shannon's work pointed out the upper bound of the transmission rate of a given communication system which is also valid in mathematical sense.

2.2   Information Theory Problems in Cryptology

Cryptology is another topic discussed in the paper. Shannon used two functions to described the operations performed by the transmitter and receiver in encoding

and decoding the information that:

$$y_n = f(x_n, \alpha_n)$$

$$\alpha_{n+1} = g(x_n, \alpha_n)$$

where x_n is the nth input symbol, α_n is the state of the transducer when the nth input symbol is introduced, and y_n is the output symbol produced when x_n is introduced if the state is α_n.

2.3   Information Theory Problems in Coding Methods

As discussed before, coding methods with high efficiency will enhance the transmission rate of different channels. In order to obtain a transmission rate approximated to the theoretical maximum channel capacity, a practical problem in information theory is how to design an efficient coding method such that the transmission rate is as higher as possible.

Two forms of efficient coding discussed in the paper are data compression codes and error-detecting codes.

2.4   Information Theory Problems in Linguistic

In Shannon's paper, information theory also provides a means for measuring redundancy of symbolic representation within a given language. For instance, English as a language has a measurable redundancy of 50%.

The redundancy can be obtained both from calculations of entropy, and also attempts aiming to recover a arbitrarily in-completed text written in a given language.

3   Inspect to Details of The Paper

One Mathematical Gap

In the part of defining the capacity of a discrete channel, Shannon gave an example of teletype that:

It is easily seen that in the teletype case this reduces to the previous result. It can be shown that the limit in question will exist as a finite number in most cases of interest. Suppose all sequence of the symbols $S_1, ..., S_n$ are allowed and these symbols have durations $t_1, ..., t_n$. What is the channel capacity? If $N(t)$ represents the number of sequences of duration t we have

$$N(t) = N(t - t_1) + N(t - t_2) + ... + N(t - t_n)$$

The total number is equal to the sum of the numbers of sequences ending in $S_1, S_2, ..., S_n$ and these are $N(t - t_1), N(t - t_2), ..., N(t - t_n)$, respectively. According to a well-known result in finite differences, $N(t)$ is then asymptotic for large t to X_0^t where X_0 is the largest real solution of the characteristic equation:

$$X^{-t_1} + X^{-t_2} + ... + X^{-t_n} = 1$$

and therefore

$$C = log X_0$$

Actually the equivalence of $N(t) = N(t - t_1) + N(t - t_2) + ... + N(t - t_n)$ will implicitly regard the total time t as being **fully occupied** by n sorts of symbols. However, let's assume a particular case that there are two different symbols, S_1 and S_2, and if the duration of the first symbol $d_1 = 2$, and the duration of the second

symbol $d_2 = 4$, then any summed time period of these two symbols are even. If the total time t is an odd number, then there's no such $N(t - d_1)$ and $N(t - d_2)$ that $N(t) = N(t - d_1) + N(t - d_2)$.

It is true that the reminder of the accumulative sum would become more unimportant if we are dealing with a extremely long time, but the derivation of the equivalence is not sufficient enough.