Case Study of Shannon Fano Coding

Bachelor of Technology
In
Electronics & Communication Engineering

By

NEKKANTI GUNA SAI KIRAN

ENROLLMENT No. : BT20ECE075

Under the Guidance of
Dr. Rashmi Pandhare

Indian Institute of Information Technology, Nagpur

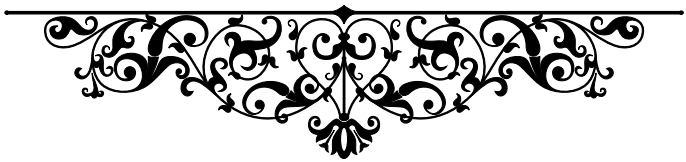
Coding Techniques

I

Details

- **Name :** Nekkanti Guna Sai Kiran
- **Enrollment No :** BT20ECE075
- **Lab Experiment:** 04
- **Course Title :**Coding Techniques
- **Lab Name :**Case Study On Shannon fano Coding In Information theory
- **Date :**16/02/2023

Case Study on Shannon Fano Coding



I

Introduction to Lossless data Compression

2.1.1 Information Theory And Source Coding

The most significant feature of the communication system is its unpredictability or uncertainty. The transmitter transmits at random any one of the pre-specified messages. The probability of transmitting each individual message is known. Thus our quest for an amount of information is virtually a search for a parameter associated with a probability scheme. The parameter should indicate a relative measure of uncertainty relevant to the occurrence of each message in the message ensemble.

The principle of improbability (which is one of the basic principles of the media world)–"if a dog bites a man, it's no news, but if a man bites a dog, it's a news"—helps us in this regard.

Hence there should be some sort of inverse relationship between the probability of an event and the amount of information associated with it. The more the probability of an event, the less is the amount of information associated with it, vice versa.

I(xj)=f(1/p(Xj)),
Where Xj is an event with a probability p(Xj) the amount of information associated with it is I(xj).
Now let there be another event Yk such that Xj Yk are independent . Hence probability of the joint event is p(Xj, Yk)= p(Xj) p(Yk) with associated information content ,

I(Xj, Yk)=f(1/p(Xj, Yk))= f(1/p(Xj) (Yk))

The total information I(Xj, Yk) must be equal to the sum of individual information I(Xj) I(Yk), where I(Yk)=f(1/p(Yk)).

Thus it can be seen that function f must be a function which converts the operation of multiplication into addition.

LOGARITHM

It is one such function. Thus, the basic equation defining the amount of information (or self-information) is, I(Xj) = log(1/P(Xi)= - log(P(Xi)
when base is 2 (or not mentioned)the unit is bit, when base is e the unit is nat, when base is 10 the unit is decit or Hertley.

ENTROPY:

Entropy is defined as the average information per individual message.
Let there be L different messages m1, m2…….mL, with their respective probabilities of occurrences be p1, p2…….pL. Let us assume that in a long time interval, M messages have been generated. Let M be very large so that M»L The total amount of information in all M messages.

The number of messages m1=M*p1, the amount of information

in message m1= log2(1/P(Xi)), thus the total amount of information in all m1 messages = M*p1* log2(1/P(Xi)).

So, the total amount of information in all L messages will then be I= M*(P1)*log(1/P1)+ M*(P2)*log(1/P2) +……+ M*(PL)*log(1/PL)

So, the average information per message, or entropy, will then be H=I/M=(P1)*log(1/P1)+ (P2)*log(1/P2) + ……+ (PL)*log(1/PL);
Hence, H(X) = - P(Xi) log2(P(Xi)) , summation i=1 to L

The Entropy of a source in bits/symbol is given by H(X) = - P(Xi) log2(P(Xi)) log2L , summation i=1 to L Where Xi are the symbols with probabilities P(Xi) , i=1,2,3 L
The equality holds when the symbols are equally likely.

There are two types of code possible:
1) Fixed Length Code : All code words are of equal length
2) Variable Length Code: All code words are not of equal length. In such cases, it is important for the formation of uniquely decodable code that all the code words satisfy the PREFIX CONDITION, which states that "no code word forms the prefix of any other code word".

The necessary sufficient condition for the existence of a binary code with code words having lengths n1 n2 ..... nL that satisfy the prefix condition is,
2power(-nk) 1 , summation k = 1… ..... L ,

Source Coding Theorem:

Let X be ensemble of letters from a discrete memory less source with finite Entropy H(X) output symbols Xi with probabilities P(Xi) , i=1,2,3… L
It is possible to construct a code that satisfies the prefix condition has an average length R that satisfies the following inequality,

H(x) R < H(x)+1, the efficiency of the prefix code is defined as
= H(x)/R,
where, H(X) = - P(Xi) log2(P(Xi)) ,summation i=1 to L

R = - ni log2(P(Xi)) summation i=1 to L

Here ni denotes the length of ith code word

The source coding theorem tells us that for any prefix code used to represent the symbols from a source, the minimum number of bits required to represent the source symbols on an average must be at least equal to the entropy of the source. If we have found a prefix code that satisfies R=H(x) for a certain source X, we must abandon further search because we can not do any better. The theorem also tells us that a source with higher entropy (uncertainty) requires on an average, more number of bits to represent the source symbols in terms of a prefix code.

Proof:

Lower bound:
First consider the lower bound of the inequality. For codewords that have length nk , 1 k L, the difference H(x)-R can be expressed as

- H(x)-R = P(Xk) log2(1/P(Xk)) - P(Xk) nk , summation K=1 L
- H(x)-R = P(Xk) log2(2power(- nk)/P(Xk)) , summation k=1 L
- H(x)-R (log2e) P(Xk) (2power(- nk)/P(Xk))-1 , summation k=1.....L
- H(x)-R (log2e) (2power(- nk) – 1 , summation k=1 L
- H(x)-R 0 (using KRAFT'S INEQUALITY)
- H(x) R

Upper bound:

Let us select a code word length nk such that

2power(- nk) P(Xk) < 2power(- nk + 1)

First consider, 2power(- nk) P(Xk)

- 2power(-nk) P(Xk) = 1 , summation k=1 L
- log2(P(Xk)) < (- nk + 1)
- nk < 1 - log2(P(Xk))
- P(Xk) nk < P(Xk) + P(Xk) log2(P(Xk)) , summation k=1 L
- R < H(x)+1

II

Shannon Fano Coding

2.2.1 Shannon Fano Coding Algorithm

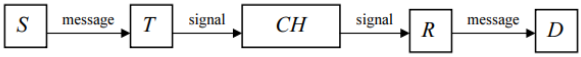
Shannon coding is yet another coding algorithm, which follows PREFIX Condition can achieve compression performance close to that of Huffman coding.

With his paper "The Mathematical Theory of Communication" (1948), Shannon offered precise results about the resources needed for optimal coding and for error-free communication. This 3 paper was immediately followed by many works of application to fields as radio, television and telephony. Shannon's theory was later mathematically axiomatized (Khinchin 1957).

According to Shannon (1948; see also Shannon and Weaver 1949), a general communication system consists of five parts:

- A **source S**, which generates the message to be received at the destination.
- A **transmitter T**, which turns the message generated at the source into a signal to be transmitted. In the cases in which the information is encoded, encoding is also implemented by this system.
- A **channel CH**, that is, the medium used to transmit the signal from the transmitter to the receiver.

- A **receiver R**, which reconstructs the message from the signal.
- A **destination D**, which receives the message.



The source S is a system with a range of possible states S1 ,..., Sn usually called letters, whose respective probabilities of occurrence are p(s1),...p(sn).
The amount of information generated at the source by the occurrence of Si is defined as:

$$I(s_i) = \log(1/p(s_i)) = -\log p(s_i) \quad (1)$$

Since S produces sequences of states, usually called *messages*, the *entropy of the source S* is defined as the average amount of information produced at the source:

$$H(S) = \sum_{i=1}^{n} p(s_i)\log(1/p(s_i)) = -\sum_{i=1}^{n} p(s_i)\log p(s_i) \quad (2)$$

Analogously, the destination D is a system with a range of possible states $d_1,...,d_m$, with respective probabilities $p(d_1),...,p(d_m)$. The amount of information $I(d_j)$ received at the destination by the occurrence of d_j is defined as:

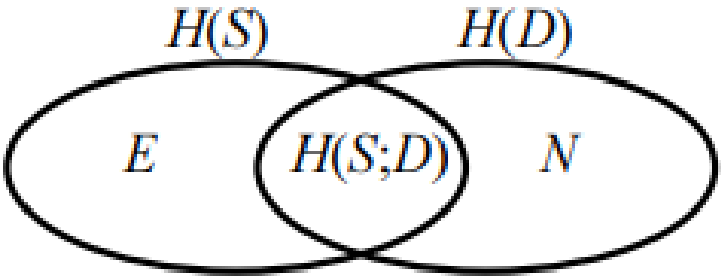
$$I(d_j) = \log(1/p(d_j)) = -\log p(d_j) \quad (3)$$

And the entropy of the destination D is defined as the average amount of information received at the destination:

$$H(D) = \sum_{j=1}^{m} p(d_j)\log(1/p(d_j)) = -\sum_{j=1}^{m} p(d_j)\log p(d_j)$$

In his original paper, Shannon (1948, p. 349) explains the convenience of the use of a logarithmic function in the definition of the entropies: it is practically useful because many important parameters in engineering vary linearly with the logarithm of the number of possibilities; it is intuitive because we use to measure magnitudes by linear comparison with unities of measurement; it is mathematically more suitable because many limiting operations in terms of the logarithm are simpler than in terms of the number of possibilities. In turn, the choice of a logarithmic base amounts to a choice of a unit for measuring information. If the base 2 is used, the resulting unit is called 'bit' –a contraction of binary unit. With these definitions, one bit is the amount of information obtained when one of two equally likely alternatives is specified.

The relationship between the entropies of the source H(S) and of the destination H(D) can be represented in the following diagram:



where,

- **H(S;D)**is the mutual information: the average amount of information generated at the source S and received at the destination D

- **E** is the equivocation: the average amount of information generated at S but not received at D.

- **N**, is the noise: the average amount of information received at D but not generated at S,
  As the diagram clearly shows, the mutual information can be computed as:
  H(S;D)=H(S)-E=H(D)-N
  Equivocation E and noise N are measures of the dependence between the source S and the destination D

- If S and D are completely independent, the values of E and N are maximum (E=H(S) and N=H(D), and the value of H(S;D) is minimum (H(S;D) =0).

- If the dependence between S and D is maximum, the values of E and N are minimum (E=N=0), and the value of H(S;D)is maximum (H(S;D)=H(S)=H(D)) .

The values of E and N are functions not only of the source and the destination, but also of the communication channel CH. The introduction of the communication channel leads directly to the possibility of errors in the process of transmission: the channel CH is defined by the matrix

where P(d/s) is the conditional probability of the occurrence of d in the destination D given that s occurred in the source S, and the elements in any row add up to 1. On this basis, E and N can be computed as:

$$N = \sum_{i=1}^{n} p(s_i) \sum_{j=1}^{m} p(d_j/s_i) \log(1/p(d_j/s_i))$$

$$E = \sum_{j=1}^{m} p(d_j) \sum_{i=1}^{n} p(s_i/d_j) \log(1/p(s_i/d_j))$$

where $p(s_i/d_j) = p(d_j/s_i)p(s_i)/p(d_j)$. The *channel capacity* C is defined as:

$$C = \max_{p(s_i)} H(S;D)$$

where the maximum is taken over all the possible distributions p(s) at the source. C is the largest average amount of information that can be transmitted over the communication channel CH

On the other hand, in the early 1940s, it was thought that the increase of the rate in the information transmission over a communication channel would always increase the probability of error. The Second Theorem, or Noisy-Channel Coding Theorem, surprised the communication theory community by proving that that assumption was not true as long as the communication rate was maintained below the channel capacity. The channel capacity is equal to the maximum rate at which the information can be sent over the channel and recovered at the destination with a vanishingly low probability of error.

The formal simplicity of Shannon's theory might suggest that the interpretation of the involved concepts raises no difficulty. As we will see in the following sections, this is not the case at all.

III

Algorithm

Shannon coding is yet another coding algorithm, which follows PREFIX Condition can achieve compression performance close to that of Huffman coding.
Following are the steps to obtain the codewords by using Shannon Coding Algorithm:

- Arrange the symbols along with their Probability in decreasing order of Probabilities.
- Select all those symbols on one side whose probabilities is closest to half of the sum of probabilities of all symbols, remaining symbols on the other side assign the values 0 1 (first bit of their codeword) respectively to all the symbols in each of the two groups.
- Select the group having 0 assigned to each of its symbol in step 2 repeat step 2 for this group. Repeat the same task for the other group (whose symbols have been assigned 1 in the step 2)
- Repeat step 3 till a group remains to exist like above having more than one symbol.

Symbol	Stage I	Step I	Step II	Step III	Step IV	Step V
So	0.30	0	0			00
S1	0.25	0	1			01
S2	0.20	1	0			10
S3	0.12	1	1	0		110
S4	0.08	1	1	1	0	1110
S5	0.05	1	1	1	1	1111

$$H(X) = 2.36 \text{ b/symbol}$$
$$L = 2.38 \text{ b/symbol}$$
$$\eta = H(X)/L = 0.99$$

The redundancy γ is defined as

$$\gamma = 1 - \eta$$

IV

Conclusion

Despite of its formal precision and its great many applications, Shannon's theory still offers an active terrain of debate when the interpretation of its main concepts is the task at issue. In this article we have tried to analyze certain points that still remain obscure or matter of discussion, and whose elucidation contribute to the assessment of the different interpretative proposals about the concept of information. Moreover, the present argumentation might shed light on the problems related with the so-called 'quantum information theory', in particular as formulated by Benjamin Schumacher (1995) on the basis of an explicit analogy with the first Shannon coding theorem. Furthermore, the discussion about the interpretation of the concepts involved in Shannon's theory would turn out to be particularly relevant if, as some believe, there were not two kinds of information classical and quantum, but only information encoded in different ways

V

ACKNOWLEDGMENT

This work was supported by Indian Institute of Information Technology Nagpur to promote research and innovation on Shannon Fano Coding Under the guidance of Dr. Rashmi Pandhare

VI

REFERENCES

- Aczel, J. and Forte, B. (1986), " Generalized entropies and the maximum entropy principle, In: Bayesian Entropy and Bayesian Methods in Applied Statistics, edited by J. H. Justice, Cambridge University Press, Cambridge, pp. 95-100.
- Bar-Hillel, Y. and Carnap, R. (1952), "An outline of a theory of semantic information," Tech. Rep. No., 247, Research Lab. of Electronics, MIT.
- Berger, T. (1971), Rate Distortion Theory, Englewood Cliffs, N.J.: Prentice-Hall.
- Brillouin, L. (1962) Science and Information Theory, Academic Press, New York.