Information Theory

In 1948, Claude Shannon published a paper called 'A mathematical Theory of communication'. This Paper heralded a transformation in our understanding of information. Before Shannon's paper, information had been viewed as a kind of poorly defined miasmic fluid. But after Shannon's paper, it become apparent that information is a well defined and, above all measurable quantity.

"A basic idea in information theory is that information can be treated very much like a physical quantity such as energy or mass."

~ Claude Shannon, 1985

Information theory defines definite, unreachable limits on precisely how much information can be communicated between any two components of any system, whether the system is man made or natural.

Problem:

Symbol and probability (A part of my name 'UDDIN'):

Symbol (x)	U	D	I	N
Probability $P(x)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

Find the information content of each symbol, entropy of source.

Solution:

We know,

Information content, $I(x_i) = \log_2\left(\frac{1}{P(x_i)}\right)$

Message 'U':

$$I(U) = \log_2 \frac{1}{P(U)} = \log_2 \frac{1}{\frac{1}{5}} = 2.322 \text{ bits}$$

Message 'D':

$$I(D) = \log_2 \frac{1}{P(D)} = 1.322 \text{ bits}$$

Message 'I':

$$I(I) = \log_2 \frac{1}{P(I)} = 2.322 \text{ bits}$$

Message 'N':

$$I(N) = \log_2\left(\frac{1}{P(N)}\right) = 2.322 \text{ bits}$$

Again,

We know,

The entropy of the source. $H(X) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$

$$\therefore H(X) = -\sum_{i=1}^{4} P(x_i) \log_2 P(x_i)$$

$$= \frac{1}{5} \times 2.322 + \frac{2}{5} \times 1.322 + \frac{1}{5} \times 2.322$$

$$+ \frac{1}{5} \times 2.322$$

$$= 4.6982 \text{ bits/symbol.}$$

Shannon-Fano Coding

Shannon-Fano coding, named after Claude Elwood Shannon and Robert Fano, is a technique for constructing a prefix code based on a set of symbols and their probabilities.

The algorithm works and it produces fairly efficient variable-length encodings; when the two smaller sets produced by a partitioning are in fact of equal probability, the one bit of information used to distinguish them is used most efficiently. Unfortunately, shannon-Fano does not always produce optimal prefix codes.

Problem 01:- Design Shannon-Fano for the following text 'BORHAN'. Find code efficiency and redundancy.

Solution:

Number of total symbols, $N = 6$

Probability of each symbol,

$P(B) = \frac{1}{6}$, $P(O) = \frac{1}{6}$, $P(R) = \frac{1}{6}$

$P(H) = \frac{1}{6}$, $P(A) = \frac{1}{6}$, $P(N) = \frac{1}{6}$

Encoding the source symbols using Shannon-Fano encoder gives :-

Symbol	Probability	Code			Length (l_i)
B	$\frac{1}{6}$	0	0		2
O	1/6	0	1	0	3
R	1/6	0	1	1	3
H	1/6	1	0		2
A	1/6	1	1	0	3
N	1/6	1	1	1	3

Code Table

The Entropy of the source,

$$H = -\sum_{i=0}^{5} P_i \log_2 P_i = 2.58 \text{ bits/symbol}$$

The average length ~~length~~ of the binary code is,

$$L_{avg} = \sum_{i=0}^{5} P_i l_i = 2.66 \text{ bit/symbol.}$$

Thus, the code efficiency,

$$\eta = \frac{H}{L_{avg}} \times 100 = \cancel{98}.\ 96.99\%$$

$$\text{Redundancy} = 100 - \eta = 100 - 96.99$$

$$= 3.01\%.$$

Code tree: