A study on minimum Redundancy achived by Huffman Codes

Dr. Ashwani Kumar, Assistant Professor, Seth G.L. Bihani S.D.(P.G.) College, Sri Ganganagar
Parveen Single, Assistant Professor, Seth G.L. Bihani S.D.(P.G.) College, Sri Ganganagar

ABSTRACT

It has been recently proved that the redundancy r of any discrete memory less source satisfies $r \le 1 - H(P_N)$, Where P_N is the least likely source letter probability.

This bound is achieved only by sources consisting of two letters. We prove a sharper boun if the number of source letters is greater than two. Also provided is a new upper bound on r, in terms of the two least likely source letter probabilities, which improves on a previous results.

1. INTRODUCTION

Let C = {$x_1, x_2, x_3, \ldots\ldots, x_n$} be a code for source 'S' and let $n_1 \le n_2 \le n_3 \ldots\ldots \le n_N$ be the code word lengths withour the loss of generality we assume that $p_1 \ge p_2 \ge p_3 \ldots\ldots \ge p_N$. The Huffman encoding algorithm [1952] defined as the difference between the average code word lengths 'L' and the entropy H(p_1 p_2, p_3, ……p_N) of the source.

$$r = L - H(p_1, p_2, p_3, \ldots\ldots, p_N) = \sum_{i=1}^{N} p_i n_i + \sum_{r=1}^{N} p_i \log p_i$$

Capocelli and Santis [1989] who proved that as a function of P_N, the redundancy 'r' of Huffman codes is upper bounded buy

$$r \le 1 - H(P_N) \quad \ldots..1.1$$

Where H is the binary entropy function and

$$H(p) = -p \log p - (1-p)\log(1-p) \quad \ldots.1.2$$

We prove that for $N \ge 3$ the following bound, in terms of the least likely source letter probability, holds :

$$r \le \begin{cases} 1 - H(2p_N) & \text{if } 0 < p_N \le \delta \\ 0.5 + 1.5p_N - H(P_N) & \text{if } \delta < p_n \le 1/3 \end{cases} \quad \ldots.1.3$$

where d=01525. This bound is the bset possible expressed only in terms of p_N for every $p_N > 0$ and $N \ge 3$.

Prisco and Santis [1995] defined redundancy 'r' of a source, whose most and least likely source probabilities are respectively p_i and p_N, is upper bounded by

$$r \le p_1 + 0.086 - p_N \quad for\ 0 < p_1 \le 1/6 \quad \ldots1.4$$

$$r \le 2 - 1.3219(1-p_1) - H(p_1) - p_N \quad for\ 1/6 < p_1 \le 0.1971 \quad \ldots1.5$$

$$r \le 4 - 18.609\, p_1 - H(5p_1) - p_N \quad for\ 0.1971 < p_1 \le 0.2 \quad \ldots1.6$$

$$r \le 2 - 1.25(1-p_1) - H(p_1) - p_N \quad for\ 0.2 < p_1 \le 0.3138 \quad \ldots1.7$$

$$r \le 3 - (3 + 3\log 3)\, p_1 - H(3p_1) - p_N \quad for\ 0.3138 < p_1 \le 1/3 \quad \ldots1.8$$

$$r \le 1 - 0.5(1-p_1) - H(p_1) - 2p_N \quad for\ 1/3 < p_1 \le 0.4505, N \ge 6 \quad \ldots1.9$$

2. Redundancy of Huffman codes:

Now we define a function of lease likely source letter probability when N=425 is the form of following theorems are te3adious case by case proof indeed we distinguish among all possible length vectors of the Huffman codes and then we proceed in the way similar to Prisco and Santis.

Theorem 2.1

Let S = (P_1, P_2, P_3 P_4) be a discrete source and $p_4 = p_N$ be its least likely source letter probability .The redundancy of the corresponding Huffman code is upper bounded by

$$
\begin{cases}
1+5p_N - H(1-3p_N, p_N, p_N, p_N) & if\ 0 < p_N \le 1/9 \\
2 - H\left(1/3, 1/3, 1/3 - p_N, 1/3\right) & if\ 1/9 < p_N \le 1/6 \\
2 - H(2p_N, 1-4p_N, p_N, p_N) & if\ 1/6 < p_N \le \delta_1 \\
2 - H\left(\frac{1+p_N}{3}, \frac{1-2p_N}{3}, \frac{1-2p_N}{3}, p_N\right) & if\ \delta_1 < p_N \le 1/5 \\
2 - H(1-3p_N, p_N, p_N, p_N) & if\ 1/5 < p_N \le 1/4
\end{cases}
$$

where $\delta_1 = 0.1708$ is the unique point in the interval $\left[1/6, 1/5\right]$ for which function H(2x,1-4x,x,x) is equal to the function $H\left[\left(1+x/3\right),\left(1-2x/3\right),\left(1-2x/3\right),x\right]$. The bound is tight.

Theorem 2.2

Let S = (P_1, P_2, P_3, P_4,P_5) be a discrete source and $P_5 = P_N$ be its least likely source letter probability. The redundancy of the corresponding Huffman code is upper bounded by

$$
\begin{cases}
1+8p_N - H(1-4p_N, p_N, p_N p_N) & if\ 0 < p_N \le \delta_2 \\
13/6 + p_N/2 - H\left(1/3, 1/3, \frac{1-3p_n}{6}, \frac{1-3p_n}{6}, p_N\right) & if\ \delta_2 < p_N \le 1/9 \\
2+2p_N - H(3p_N, 1-6p_N, p_N, p_N, p_N) & if\ 1/9 < p_N \le 1/8 \\
2+2p_N - H\left(\frac{1-2p_N}{6}, \frac{1-4p_N}{3}, p_N, p_N, p_N\right) & if\ 1/8 < p_N \le \delta_3 \\
\frac{11}{5} + \frac{4p_N}{5} - H\left(\frac{2(1-p_N)}{5}, \frac{1-p_N}{5}, \frac{1-p_N}{5}, p_N\right) & if\ \delta_3 < p_N \le 1/6 \\
2+2p_N - H(1-4p_N, p_N, p_N, p_N, p_N) & if\ 1/6 < p_N \le \delta_4 \\
\frac{9}{4} + \frac{3p_N}{4} - H\left(\frac{1-p_N}{4}, \frac{1-p_N}{4}, \frac{1-p_N}{4}, \frac{1-p_N}{4}, p_N\right) & if\ \delta_4 < p_N \le 1/5
\end{cases}
\quad \ldots 2.2.1
$$

where $\delta_1 = 0.078184$ is the unique point in $\left]0, 1/9\right[$ for which the function 1+8x-H(1-4x,x,x,x,x) is equal to the function $13/6 + x/2 - H\left(1/3, 1/3, \frac{1-3x}{6}, \frac{1-3x}{6}, x\right)$; $\delta_1 = 0.143815$ is the unique point in $\left]1/8, 1/6\right[$ for which the function $2+2x-H\left(\frac{(1-2x)}{2}, \frac{(1-4x)}{2}, x, x, x\right)$ is equal to the function $\frac{11}{5} + \frac{4x}{5} - H\left(\frac{2(1-x)}{5}, \frac{1-x}{5}, \frac{1-x}{5}, x\right)$ and $\delta_4 = 0.179669$ is the unique point in

$\left]\frac{1}{6}, \frac{1}{5}\right[$ for which the function 2+2x-H(1-4x,x,x,x) is equal to the function $\frac{9}{4} + \frac{3x}{4} - H\left(1 - \frac{x}{4}, 1 - \frac{x}{4}, 1 - \frac{x}{4}, 1 - \frac{x}{4}, x\right)$. The bound is tight.

3. Latest Upper Bound of Redundancy :

We give the least upper bound as a function of least likely source letter probability when N=4 & 5. We distinguish among all possible length vectors of Huffman codes and then we proceed in a similar way to presco and Sanits and define as

Theorem 3.1

Let S = (P_1, P_2, P_3, P_4) be a discrete source and $p_4 = p_x$ be its least likely source letter probability. The redundancy of the corresponding Huffman code is upper bounded by

$$r = \begin{cases} 1 + 4p_N - H(1 - 3p_N, p_N, p_N, p_N) & \text{if } 0 < p_N \le \frac{1}{9} \\ 1.96 - H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2} - p_N, p_N\right) & \text{if } \frac{1}{9} < p_N \le \frac{1}{6} \\ 1.81059 - H(3p_N, 1 - 5p_N, p_N, p_N) & \text{if } \frac{1}{6} < p_N \le \delta_1 \\ 1.98768 - H\left(\frac{1+2p_N}{3}, \frac{1-2p_N}{3}, \frac{1-p_N}{3}, p_N\right) & \text{if } \delta_1 < p_N \le \frac{1}{5} \\ 2 - H(1 - 3p_N, p_N, p_N p_N) & \text{if } \frac{1}{5} < p_N \le \frac{1}{4} \end{cases}$$

where $\delta_1 = 0.1708$ is the unique point in the interval $\left]\frac{1}{6}, \frac{1}{5}\right[$ for which the function 1.81059-H(3x,1-5x,x,x) is equal to the function $1.98768 - H\left(\frac{1+2x}{3}, \frac{1-2x}{3}, \frac{1-x}{3}, x\right)$. The bound is tight.

Theorem 3.2

Let S=(P_1, P_2,P_3, P_4, P_5) be a discrete source and $P_5 = P_N$ be its least likely source letter probability The redundancy of the corresponding Huffman code is upper bounded by

$$r \le \begin{cases} 1.0228 + 3p_N - H\left(1 - 3p_N, \frac{p_N}{2}, \frac{p_N}{2}, p_N, p_N\right) & \text{if } 0 < p_N \le \delta_2 \\ 2.09563 + \frac{p_N}{3} - H\left(\frac{1}{3}, \frac{1}{3}, \frac{1-2p_N}{6}, \frac{1-4p_N}{6}, p_N\right) & \text{if } \delta_2 < p_N \le \frac{1}{9} \\ 1.953 + p_N - H(4p_N, 1 - 7p_N, p_N, p_N, p_N) & \text{if } \frac{1}{9} < p_N \le \frac{1}{8} \\ 2.20742 + \frac{p_N}{25} - H\left(\frac{1-3p_N}{3}, \frac{2-6p_N}{3}, p_N, p_N, p_N\right) & \text{if } \frac{1}{8} < p_N \le \delta_3 \\ \frac{11}{5} + \frac{1.1p_N}{6} - H\left(\frac{2(1-p_N)}{5}\right), \frac{1-2p_N}{5}, \frac{1-p_N}{5}, \frac{1-p_N}{5}, p_N & \text{if } \delta_3 < p_N \le \frac{1}{6} \\ 1.967 + \frac{p_N}{2.2} - H\left(1 - 3p_N, \frac{p_N}{2}, \frac{p_N}{2}, p_N, p_N\right) & \text{if } \frac{1}{6} < p_N \le \delta_4 \\ 2.107547 + \frac{p_N}{2} - H\left(\frac{1-2p_N}{6}, \frac{1-2p_N}{6}, \frac{2-p_N}{6}, \frac{2-p_N}{6}, p_N\right) & \text{if } \delta_4 < p_N \le \frac{1}{5} \end{cases}$$

Where δ_2=0.078184 is the unique point $\left]0, \frac{1}{9}\right[$ for which the function $1.0228+3x-H\left(1-3x, \frac{x}{2}, \frac{x}{2}, x, x\right)$ is equal to the function $2.09563+\frac{x}{3}-H\left(\frac{1}{3}, \frac{1}{3}, \frac{1-2x}{6}, \frac{1-4x}{6}, x\right)$, δ_3=0.143815 is the unique point in $\left]\frac{1}{8}, \frac{1}{6}\right[$ for which the function $2.20742+\frac{x}{25}-H\left(\frac{1-3x}{3}, \frac{2-6x}{3}, x, x, x\right)$ is equal to the function $\frac{11}{5}+\frac{1.1x}{6}-H\left(\frac{2(1-x)}{5}, \frac{1-2x}{5}, \frac{1-x}{5}, \frac{1-x}{5}, x\right)$ and δ_4=0.179669 is the unique point in $\left]\frac{1}{6}, \frac{1}{5}\right[$ for which the function $1.967+\frac{x}{2.2}-H\left(1-3x, \frac{x}{2}, \frac{x}{2}, x, x\right)$ is equal to the function $2.09563+\frac{x}{3}-H\left(\frac{1}{3}, \frac{1}{3}, \frac{1-2x}{6}, \frac{1-4x}{6}, x\right)$ the bound is tight.

4. Conclusion : - The above result is obtain by a study on minimum Redundancy achived by Huffman codes.

References :-

(i) Abramson. N(1963): Information theory and coding. McGraw Hill, Mew York.

(ii) AczeL J. (1966): Lectures on functional equations and their applications. Academic Press New York.

(iii) AczeL L (1968): On different characterizations of entropies. Proc. Internal. Symp. Probability and Information Theory. McMaster University Lecture Notes in Mathematics, Springer-Verlag, Vol. 89, 1-11

(iv) AczeL L. (1970): A mixed theory of information II. additive inset entropies of randomized systems of events with measurable sum property. Utilitas Mathematica, 13, 49-54.

(v) AczeL L. (1980): A mixed theory of information VII, Inset information function pf degrees. C.R. Math Rep. Acad. Sci. Canada 2,125-129

(vi) AczeL L. and B. Forte (1970): A system of axioms for the measurfe of the uncertainty. Notices Amer. Math. Soc. 17. 202.

(vii) AczeL L. and Kaimappan (1978): A mixed theory of Information II Inset entropies of degree b. Information and Control 39, 315-322.