Shannon-Fano Code

September 19, 2013

Example

Given a set of symbols and their probabilities of occurrence

\[ S = \{a, b, c, d, e\} \quad P(S) = \{0.4, 0.3, 0.15, 0.1, 0.05\} \]

First find the entropy (average information content per message)

\[ H = \sum p_i I_i = \sum p_i \log_2 \frac{1}{p_i} = 2.0087 \text{ bit/symbol} \]

If normal binary code is used

\[
\begin{align*}
  a & : 000 \\
  b & : 001 \\
  c & : 010 \\
  d & : 011 \\
  e & : 100
\end{align*}
\]

Average code length \(3\)

Therefore, we use a 3 bit code word to transmit 2.0087 bit of information (on average), the efficiency and redundancy are

\[ \eta = \frac{H}{L_{avg}} = \frac{2.0087}{3} = 67\% \quad \gamma = 1 - \eta = 33\% \]

Now consider Shannon-Fano Code

The idea of Shannon-Fano Code is to first group the symbol into 2 group with equal probabilities (or as close as possible)

\[ a_{0.4}, b_{0.3}, c_{0.15}, d_{0.1}, e_{0.05} \]

The grouping is

\[ \begin{array}{c}
  a_{0.4}, d_{0.1} \\
  0.5 \\
  \underline{b_{0.3}, c_{0.15}, e_{0.05}} \\
  0.5
\end{array} \]

Then assign the first bit

\[ a_{0.4}(0), d_{0.1}(0) \quad b_{0.3}(1), c_{0.15}(1), e_{0.05}(1) \]

Repeat the step, the new grouping is

\[ \begin{array}{c}
  a_{0.4}(0) \quad d_{0.1}(0) \\
  0.5 \\
  b_{0.3}(1) \quad c_{0.15}(1), e_{0.05}(1)
\end{array} \]

Assign next bit (notice that although \(b\) is separated, it still need to assign one more bit because of pre-fix code condition !)

\[ a_{0.4}(00) \quad d_{0.1}(01) \quad b_{0.3}(10) \quad c_{0.15}(11), e_{0.05}(11) \]
Lastly split c and e

\[ a_{0.4}(00) \quad d_{0.1}(01) \quad b_{0.3}(10) \quad c_{0.15}(110) \quad e_{0.05}(111) \]

Done

The average code length is thus

\[ L_{av} = \sum p_il_i = 0.4(2) + 0.1(2) + 0.3(2) + 0.15(3) + 0.05(3) = 2.2 > H = 2.0087 \]

The efficiency now is

\[ \eta = \frac{H}{L_{av}} = \frac{2.0087}{2.2} = 91\% \quad \gamma = 1 - \eta = 9\% \]

Efficiency increased from 66% to 91%, improved a lot!

–END–