Shannon-Fano method

A simple data compression procedure, which involves the following steps:

1. Arrange the symbols to be transmitted according to their probabilities.
2. Divide the set of symbols into two subsets with equal or nearly equal probabilities. Assign the symbol set with **0** and the other with **1** as the first symbol of their codewords.
3. Continue this process until no symbols are left.

Example:

Symbol	Probability	Code	Length
X1	0.25	00	2
X2	0.25	01	2
X3	0.125	100	3
X4	0.125	101	3
X5	0.0625	1100	4
X6	0.0625	1101	4
X7	0.0625	1110	4
X8	0.0625	1111	4

We use 8 symbols for encoding, where the symbols have the given probabilities in the message to be sent. They are arranged in decreasing order of their frequency of occurrence, with the most frequent ones at the top. The third column shows the assigned variable-length codes, with shorter codes given to more frequent symbols.

The 8 symbols form a complete event system since they are mutually exclusive events, and the sum of their occurrence probabilities is 1.

Let's calculate the entropy of this event system:

$$ H = -\left(\frac{2}{4} \log \frac{1}{4} + \frac{2}{8} \log \frac{1}{8} + \frac{4}{16} \log \frac{1}{16}\right) = 2.75 \, \text{[bit].} $$

What is the average word length?

We can calculate the average word length with the use of this function: \(L = \sum_{i=1}^n x_i \mu_i \)

After substituting:

$$ L = \frac{2}{4} \dot 2 + \frac{2}{8} \dot 3 + \frac{4}{16} \dot 4 = 2.75 [bit]$$

The original 8 symbols (X1-X8) can be stored with 4 bits. The average length is 2.75 bits, thus the compressed messages will be \(\frac{2.75}{4} \Rightarrow 68.75% \) shorter.

From:
https://edu.iit.uni-miskolc.hu/ - **Institute of Information Science - University of Miskolc**

Permanent link:
https://edu.iit.uni-miskolc.hu/tanszek:oktatas:techcomm:shanon-fano_method?rev=1728376863

Last update: **2024/10/08 08:41**