Information Theory and Coding
Coding Techniques

م. فؤاد حمادي

2020 - 2019

Coding Techniques

Information is rarely transmitted directly to the receiver, without any processing due to the followings:

• source alphabet is different from channel one, therefore some adaptation of the source to the channel is needed (*information representation codes*).

• channel must be very efficiently used (close as possible to its full capacity), meaning that the source must be converted into an optimal one. For an efficient use of the channel - in order to minimize transmission time and/or storage space – source compression is needed (*compression codes*).

• it is also necessary to adapt the source to the channel to ensure synchronization between transmitter and receiver.

• information transmitted over a noisy channel is distorted by channel noise; this is why error detecting and correcting codes are used in error control procedures.

• information confidentiality from unauthorized persons must be provided in some applications. The need for source processing prior to transmission (or storage) is leading to the processing known as coding.

The problem of representing the source alphabet symbols S_i in term of another system of symbols (usually the binary system consisting of the two symbols 0 and 1) is the main topic of coding.

The two main problems of representation are the following:

1. How to represent the source symbols so that their representations are far apart in same suitable sense. As a result, in spite of small changes (noise) in their representation, the altered symbols can be

discovered to be wrong and even possibly corrected (i.e Error – Detecting codes & Error- correcting codes).

2. How to represent the source symbols in a minimal form for purposes of efficiency. The average code length is minimized, where $_i$ is the length of the representation of the ith symbol S_i the entropy function provides a lower bound on L ($L \geq H(x)$).

$$L= \sum_{i=1}^{N} P_i l_i$$

Average Length of Code L

Let a code transform the source symbols $x_1,x_2,...,x_N$ into the code symbols $C_1,C_2,...,C_N$. let the probabilities of the source symbols be $P_1,P_2,\ldots, P_N$ and let the length of the code words be $l_1, l_2,\ldots, l_N$, then

$$L= \sum_{i=1}^{N} P_i l_i$$

Average Information (Entropy)

In practice we are often more interested in the ***average*** information conveyed in some process than in the ***specific*** information in each event. For example, a source may produce a set of events of probability P_1, P_2, P_3, … , P_i,…, P_n. In a long sequence of n, event i will occur n*Pi times, contributing n * - log Pi bits. The average information over all events is called the entropy H, and is given by:

$$\bar{I}= \sum_{i=1}^{N} P_i l$$

$$\bar{I} = H = - \sum_{i=1}^{N} pi \log pi \qquad \text{bit\symbol}$$

In the case of letters of alphabet, the probabilities are not really all the same. The entropy is given by: H = - (P(A) log P(A) + P(B) log P(B) + … + P(Z) log P(Z)) ≈ 4 .1 b its (using standard values for P(A), P(B), etc.) The units are often given as bits/letter or bits/symbol to stress the average nature of the measure. The function H of the probability distribution P_i measures the amount of ***uncertainty, surprise, or information*** the distribution contains. The term "entropy" was deliberately chosen for the name of measure of average information, because of its similarity to entropy in thermodynamics. In ***thermodynamics***, entropy is a measure of the ***degree of disorder*** of a system, and disorder is clearly related to information.

Example1: A binary source produces a stream of 0s and 1s with probabilities P(0) = 1/8 and P(1) = 7/8 respectively, find the entropy of this source.

Sol:

$$H = -\sum_{i=1}^{N} p_i \log p_i$$

$$= - (1/8 \log 1/8 + 7/8 \log 7/8)$$

$$= -(0.125 \times -3 + 0.875 \times - 0.192)$$

$$= - (- 0.375 - 0.168) = 0.543 \text{ bits \textbackslash\ Symbol}$$

Example2: For the above source assume that the source send the below stream of letters **ADBAAEBACBAAACABDAAB**

Find

i. the amount of information each letter convey?

ii. the amount of information that the total stream of letters (message) convey?

Sol: We can see that the total number of letters in the message is equal to 20.

$P(A) = 10/20 = 0.5$

$P(B) = 5/20 = 0.25$

$P(C) = P(D) = 2/20 = 0.1$

$P(E) = 1/20 = 0.05$

i. $I_A = -\log 0.5 = 1$

$I_B = -\log 0.25 = 2$

$I_C = I_D = -\log 0.1 = 3.322$

$I_E = -\log 0.05 = 4.322$

ii. the amount of information that the message convey is :

$I_{message} = 10 \times 1 + 5 \times 2 + 2 \times 3.322 + 2 \times 3.322 + 1 \times 4.322$

$= 10 + 10 + 6.644 + 6.644 + 4.322$

$= 37.61$

Example3: The above source produces a stream of letters (A, B, C, D, and E) with the probabilities bellow

Letter	**A**	**B**	**C**	**D**	**E**
Probability	**0.5**	**0.25**	**0.1**	**0.1**	**0.05**

Find the entropy of this source.

Sol

$H = -\sum P_i \log P_i$

$= -(0.5 \log 0.5 + 0.25 \log 0.25 + 0.1 \log 0.1 + 0.1 \log 0.1 + 0.05 \log 0.05)$

$= -(0.5 * -1 + 0.25 * -2 + 2 * 0.1 * -3.322 + 0.05 * -4.322)$

$= 1.88$ bits/symbol

Or

H = I (message) / total number of letters

= 37.61 / 20

= 1.88 bits / symbols

Example4: Find and then compare between the entropies of the following three binary

sources source i **source ii** **source iii**

$$\begin{bmatrix} A1 & A2 \\ 1/256 & 255/256 \end{bmatrix} \quad \begin{bmatrix} B1 & B2 \\ 1/2 & 1/2 \end{bmatrix} \quad \begin{bmatrix} C1 & C2 \\ 7/16 & 9/16 \end{bmatrix}$$

Sol

H_i= -(1/256 log 1/256 +255/256 log 255/256) = 0.0369 bit/symbol

H_{ii}= -(1/2 log 1/2 + 1/2 log 1/2) =1 bit/ symbol

H_{iii}= - (7/16 log 7/16 + 9/16 log 9/16)= 0.989 bit/ symbol

Source ii give more information than source iii and source iii give more information than source i.

Example5: a source producing three letters A,B,C with the probabilities :

i . 1/3, 1/3, 1/3 respectively

ii. 1/2, 1/4, 1/4 respectively

Find and compare between the entropies of the two cases.

Sol

i. H= -3(1/3 log 1/3)

= log 3

= 1.58 bit/ symbol

ii. $H = -(1/2 \log 1/2 + 1/4 \log 1/4 + 1/4 \log 1/4)$

$= -(-0.5 - 0.5 - 0.5)$

$= 1.5$ bit/ symbol

source in case i give more information than the source in case ii.