Data Compression

Enrico Marchionni
enrico.marchionni@studio.unibo.it

December 10, 2024

Abstract

Data compression is intended as the practice of reducing the size of binary digital data. It could be considered as a procedure that takes a bit-stream in input and returns another bit-stream as output. The output stream may be of equal length or shorter than the input.

The key to understand data compression is to discuss the distinction between data and information. It can be said that data is how information is represented[1]. In simple terms, data can be compressed because its original representation is not the shortest possible. The goal of data compression is to reduce data by maintaining the same information.

The counterpart is that in our time data is intrinsically redundant. And this redundancy is needed. So data compression isn't only a procedure that goes from a bit-stream to another one not longer, but it requires also another procedure that regenerates the original bit-stream of data, necessary for practical use, from the previously given output bit-stream of information.

. . .

[1] ex. the number 0 can be expressed in binary as a sequence of a certain number of zeros, from 1 to ∞, and we know that calculators use at least 8 bits, let's say n (considering it as a multiple of 8), to represent an integer number. So at the end $n - 1$ bits are redundant in the 0 representation on a calculator.

Contents

Chapter 1

Information Theory

In 1948, Shannon[1], while working at the Bell Telephone Laboratories, published "A Mathematical Theory of Communication" [**AMathematicalTheoryOfCommunication**], a seminal paper that marked the birth of information theory. In that paper, Shannon defined the concept of "information" and proposed a precise way to quantify it-in his theory, the fundamental unit of information is the bit.

Moreover, this discipline plays behind the concepts of entropy, randomness and data compression, all topics that will be discussed later on.

1.1 Quantifying Information

For what concerns data compression, information of theory has developed a usable measure of the information we get from observing the occurrence of an event having probability p. Therefore information is defined in terms of the probability.

The information measure $I(p)$ has to match the following axioms (from [**AnIntroductionToInformationT**

- Information is non negative: $I(p) \geq 0$.
- If an event has probability 1, we get no information: $I(1) = 0$.
- If two independent events occur (whose probability is the product of their individual probabilities), then the information we get from observing the events is the sum of the two computed individually: $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$.
- Information measure must be continuous and monotonic (slight changes in probability should result in slight changes in information).

Considering the previous properties as axioms it can be said that: $I(p^2) = I(p \cdot p) = I(p) + I(p) = 2 \cdot I(p)$. Thus: $I(p^n) = n \cdot I(p)$ (by induction). Then: $I(p) = I(p^{(\frac{1}{m})^m}) = m \cdot I(p^{\frac{1}{m}})$, so $I(p^{\frac{1}{m}}) = \frac{1}{m} \cdot I(P)$, therefore: $I(p^{\frac{n}{m}}) = \frac{n}{m} \cdot I(p)$. In general, considering r as a real number: $I(p^a) = a \cdot I(p)$.

From this analysis it was discovered that:

$$I(p) = -\log_b p \; (= \log_b \frac{1}{p}) \tag{1.1}$$

[1] Claude Elwood Shannon (1916–2001) was an American mathematician, electrical engineer, computer scientist, cryptographer and inventor known as the "father of information theory".

Where: $p = b_1^{\log_{b_1} p}$ and therefore: $\log_{b_2} p = \log_{b_2} b_1^{\log_{b_1} p} = \log_{b_2} b_1 \cdot \log_{b_1} p$. So: $\log_{b_2} b_1$ is a constant, a scaling factor. From another point of view it is a simple change in the unit of measurement.

For this reason:

$$I(p) = -\log_2 p \tag{1.2}$$

Equation 1.2 is the same expression of Equation 1.1 where the unit of measurement is called bits (look at Table 1.1). Equation 1.1 was first introduced by Hartley[2] in 1928 trying to measure uncertainty, without talking about probability, and lately reviewed by Shannon.

Unit of measurement	Base
bit (or shannon)	2
trit	3
nat (natural unit of information)	e
hartley (or dit)	10

Table 1.1: Information units of measurement

Example 1. *Let's talk about flipping a fair coin n times. It gives us:* $-\log_2 \frac{1}{2}^n = \log_2 2^n = n \cdot \log_2 2 = n$ *bits of information. In fact a sequence of heads (coded as* 1*) and tails (coded as* 0*) could be expressed as:* $010010111\ldots$*, these are the* n *bits of information.*

[2] Ralph Vinton Lyon Hartley (1888-1970) was an American electronics researcher. He invented the Hartley oscillator and the Hartley transform, and contributed to the foundations of information theory.