Information Theory

Quan Wen

24 Oct 2018

Shannon Information

Shannon information can be formulated intuitively by playing the game "sixty-three". If you have a number in your hand, and this integer is randomly picked from 0 and 63 with equal probability. What's the smallest number of yes/no questions needed to identify an integer x between 0 and 63?

Intuitively, Six questions suffice. One reasonable strategy ask the following questions:

- $x \geq 32$?
- is x mod $32 \geq 16$?
- is x mod $16 \geq 8$?
- is x mod $8 \geq 4$?
- is x mod $4 \geq 2$?
- is x mod $2 \geq 1$?

What are the Shannon information contents of the outcomes in this example? If we assume that all values of x are equally likely, then the answers to the questions are independent and each has Shannon information content $\log_2(1/0.5) = 1$bit; the total Shannon information gained is always $\log_2 64 = 6$ bits. Furthermore, the number x that we learn from these questions is a six-bit binary number. Our questioning strategy defines a way of encoding the random variable x as a binary file.

Now let us put everything in the context of sensory encoding. Given a sensory stimulus, such as flash light with given intensity, neurons in the retina, such as biopolar cells, can faithfully respond to the intensity change. Let us denote the response amplitude r, which could be membrane potential, or the firing rate of the neuron. Just by measuring the response of the neuron, how much information we could learn about the stimulus. The answer is just given by

$$h = -\log_2 P(r)$$

where $P(r)$ is the probability of a given response amplitude.

When we write down the above formula, we are assuming that all responses have equal probability. When responses have unequal probabilities, we modify the above formula, by computing the average over all different responses. This leads to the definition of Shannon information

$$I = -\sum_r P(r) \log_2 P(r) \tag{1}$$

When we treat the response of a neuron as a continuous variable, one should then define the probability density function $p(r)$, and the Shannon information becomes

$$I = -\sum_r p(r)\Delta r \log_2 p(r)\Delta r \tag{2}$$

Furthermore,

$$\lim_{\Delta r \to 0}(I + \log_2 \Delta r) = -\int dr p(r) \log_2 p(r) \tag{3}$$

We now ask, what is the input-output relationship of a neuron, $r = f(s)$, that would maximize the information (or entropy) of the response? Let's consider the simple case, by assuming that there is a maximum response r_{max} a neuron could achieve. This could be the maximum firing rate. And the minimum response is zero. It is now your homework to show that under this condition, different responses with equal probability would maximize the entropy. Now use the normalization condition

$$\int_0^{r_{max}} p(r)dr = 1,$$

we found that $p(r) = \frac{1}{r_{max}}$. Given the probability density distribution of inputs $p(s)$, One must have $p(s)ds = p(r)dr$. As a result, $\frac{dr}{ds} = r_{max}p(s)$, and the input-output function should be proportional to the cumulative probability distribution of stimulus

$$f(s) = \int_0^s r_{max}p(s')ds' \tag{4}$$