

CS 811 Final Project

Ray Shulang Lei

200253624

Department of Computer Science

University of Regina

Regina, Saskatchewan, S4S0A2, Canada

April 24, 2011

Abstract

This project paper reviews Claude E. Shannon's information theory and the algorithmic information theory.

1 Introduction

2 Claude E. Shannon's information theory

2.1 The entropy of information

In information theory, the entropy of information means the measurement of the uncertainty of the information. The entropy H of an variable X is:

$$H(X) = E(I(X))$$

Where E is the expected value, and I is the information content of X . Let $x_i \in X$, and $p(x_i)$ be the probability mass function of X , then the entropy can be written as:

$$H(X) = \sum_{i=1}^n p(x_i) I(x_i) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

Here we assume the unit of entropy is bit, therefore the logarithm base is 2.

2.2 The fair coin tossing example

When tossing a coin, the outcomes are either head(x_h) or tail(x_t). Thus

$$X = \{x_h, x_t\}$$

Thus, before tossing the coin once, its entropy is:

$$H_{before}(X) = -(p(x_h)\log_2 p(x_h) + p(x_t)\log_2 p(x_t))$$

Assume this is a fair coin:

$$p(x_h) = 0.5, p(x_t) = 0.5$$

Therefore,

$$H_{before}(X) = -(0.5 \times \log_2 0.5 + 0.5 \times \log_2 0.5) = -(-0.5 - 0.5) = 1$$

Which means before toss the coin, the event would have 1 bit entropy.

After the coin has been tossed, let's say the result is tail, then its possible outcome now is only x_t :

$$X = \{x_t\}, p(x_t) = 1$$

And its entropy is now:

$$H_{after}(X) = -(1 \times \log_2 1 + 1 \times \log_2 1) = -(0 + 0) = 0$$

Thus, the entropy we have gained during the coin tossing is:

$$H(X) = H_{before} - H_{after} = 1 - 0 = 1$$

Which is 1 bit entropy. This means we need 1 bit entropy to store the each toss of a fair coin.

But what if the coin we toss is not fair?

2.3 The unfair coin tossing example

Assume we have an unfair coin that the chance it lands on its head is 0.8 and the chance it lands on its tail is 0.2:

$$p(x_h) = 0.8, p(x_t) = 0.2$$

Thus, before tossing the coin once, its entropy is:

$$H_{before}(X) = -(p(x_h)\log_2 p(x_h) + p(x_t)\log_2 p(x_t))$$

Which equals to:

$$\begin{aligned} H_{before}(X) &= -(0.8 \times \log_2 0.8 + 0.2 \times \log_2 0.2) \\ &= -(0.8 \times -0.321928094887362 + 0.2 \times -2.321928094887362) \\ &= -(-0.25754247590989 - 0.464385618977472) \\ &= 0.721928094887362 \end{aligned}$$

Again, the entropy after the coin has been tossed with a result of head is 0:

$$X = \{x_h\}, p(x_h) = 1$$

And its entropy is now:

$$H_{after}(X) = -(1 \times \log_2 1 + 1 \times \log_2 1) = -(0 + 0) = 0$$

Thus, the entropy we have gained during the unfair coin tossing is:

$$H(X) = H_{before} - H_{after} = 0.721928094887362 - 0 = 0.721928094887362$$

We can see that when the possibilities are not even, the entropy is 0.721928094887362 bit, which is less than 1 bit when the coin is fair.

2.4 The average information gain

The above two examples gives us how many information on average we would gain during a fair coin toss event and a unfair coin toss event.

Notice that these information gains are average information gains for each coin tosses considering both outcomes of heads and tails.

What if the average information gain for the coin toss event that the outcomes are heads only or tails only, then? Let's look at the following examples: We will look into $I(X)$ where I is the information content of X .

$$I(X) = -\log_2 p(x_i)$$

In the case of a fair coin toss with outcomes of heads only:

$$I(X) = -\log_2 p(x_h), p(x_h) = 0.5$$

$$\implies I(X) = -\log_2 \frac{1}{2} = 1$$

In the case of a fair coin toss with outcomes of tails only:

$$I(X) = -\log_2 p(x_t), p(x_t) = 0.5$$

$$\implies I(X) = -\log_2 \frac{1}{2} = 1$$

Thus 1 bit entropy is gained on average with each fair coin toss with the outcome of both head and tail.

In the case of the unfair coin toss with outcomes of heads only:

$$I(X) = -\log_2 p(x_h), p(x_h) = 0.8$$

$$\implies I(X) = -\log_2 \frac{4}{5} = 0.321928094887362$$

In the case of the unfair coin toss with outcomes of tails only:

$$I(X) = -\log_2 p(x_t), p(x_t) = 0.2$$

$$\implies I(X) = -\log_2 \frac{1}{5} = 2.321928094887362$$

When the outcome is head, we only gained 0.3 bit of information, where when the outcome is tail, we gained 2.3 bits of information. We discovered that more entropy is gained on average with each unfair coin toss event with the outcome of less common. We conclude that when the outcome of an event is less common, then it contains more information.

3 Algorithmic information theory

3.1 Definition