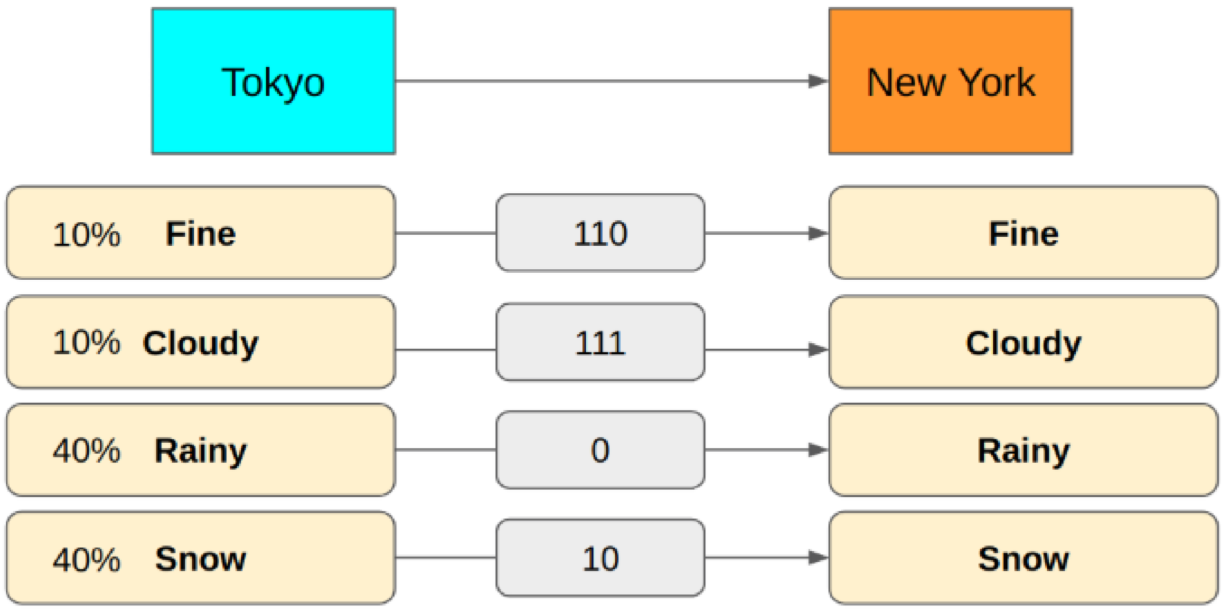
Entropy

Introduction



Average number of bits used to send messages from Tokyo to New York:

$$0.1 * 3 + 0.1 * 3 + 0.4 * 1 + 0.4 * 2 = 1.8 \, bits \; on \; average \; per \; message$$

Therefore, to obtain smaller average encoding size, **lesser number of bits** is used on case with **larger probability**.

Calculate Entropy

In general, when we need *N* different values expressed in bits, we need *n* bits:

$$n = \log_2 N$$

In other words, if a message type happens 1 out of N times, the above formula gives the minimum size required. Let **P = 1/N**:

$$n = \log_2 N = -\log_2 \frac{1}{N} = -\log_2 P$$

Minimum average encoding size (in bits) = entropy.
Recall previously, minimum average encoding size = P(i) * n

$$Entropy = -\sum_i P(i) \log_2 P(i)$$

Analogies

High entropy is associated with **disorder**, **uncertainty**, **surprise**, **unpredictability**, **more specific information**.

1. High entropy(encoding size is big on average)
   -> *many message types* / many *small prob* messages
   -> every new message arrives, expect a different type than previous message
   -> **disorder/uncertainty/unpredictability**

2. When a message types with much smaller probability than other message types happens --> **surprise**

3. A rare message type has more information than more frequent message types because it eliminates a lot of other probabilities and tells us **more specific information**.

4. High entropy
   -> many *small prob* messages
   -> each message sent have more specific information (**certain**)

Reference

[Medium]