Report of A Mathematical Theory of Communication

By Qingyuan Li 13016389

Github link: https://github.com/chris0906/UTS_ML2019_ID13016389

Introduction

Before Shannon, people don't think that information can be measured as weight, volume, and electric current. but when his paper was released out, it explained a way of measuring information. one of the most valuable contributions of Shannon is that he finds a mathematical way of measuring information which is Bit. in addition to that, he also talks about the signal to noise ratio, how could we separate signal from noise. another prominent contribution of Shannon is that he accurately calculates out how much information can an information channel contain. it's no doubt that it's a great breakthrough.

Content

Shannon gives us a measurement of a basic unit of information, which is Bit. it's defined like this: imagine there is a black box that includes A and B, let's choose one from it every time, but we don’t know which one is going to be, that means A and B has the same probability to be the next character, it’s 50% versus 50%, in order to figure out which one is going to be, we need one bit information. on the other hand, if the probability of occurrence of A is greater than B, then, we could say that we need less 1-bit information to make sure which one is the next character. Shannon describes this theory by using the mathematical formula:

H = -P1LogP1-P2LogP2-......-PnLogPn (Log is 2-based)

if we replace this formula with the corresponding number, we will get **0.5Log0.5+0.5Log0.5 = 1**.then, Shannon introduces his first law:

the length of encoding >= H(information) / H(code)

in this formula, the **H** represents how many information you want to encode, **H(code)** represents how many information each code include. as a result, we could determine how long the encoding will be. it has resolved the relationship between information and its length of the encoding. Shannon gives us an example of this:

Imagine you want to encode four-character A, B, C, D with binary encoding. so the first thing is how much information could A, B, C, D include (given each has the same probability of occurrence), based on the previous finding, we will know **H(information) = Log4 (2-based) = 2 bits.** then how many information could a binary code has, it could be either 0 or 1, which could be **H(binary) = Log2 = 1.** hence, **the length of encoding = H(information)/H(binary) = 2.** we could represent A, B, C, D as 00, 01, 10, 11. it is perfectly covering the whole information. (E.T. 1957)

Moreover, he discusses noisy and signal, and how their relationship would be when it comes to getting more information out of it, the more signal to noise ratio it is, the better information we can get. Then based on noisy discussion, Shannon introduces information channel, what's great about that is it defines how much information can an information channel pass, once the information channel can be measured by a mathematical formula, we will know how much information can be passed at most at one time in a channel. In the past, nobody defines a channel capability, once this has been done, we will know exactly what is the limitation for information transmitting.

Innovation

One of the most creative things in the paper is that no one can accurately define information unit in any mathematical way, but after this paper, the mythical of information has been revealed and based on this information measurement, he tells us how long it would be if we try to encode a sequence of information by using different code. It helps enormously in the computer world because it has to be encoded for any information in the computer, and his formula tells us how long of encoding length with different code. Interestingly in the beginning. He uses the Markov model (LR 1989) for how we can think about communication. First, he has used a zeroth-order approximation to randomly choose a letter independently from ABC to form a sequence of letters, however, the sequence doesn't look like the original one. then he uses first-order approximation, choose each letter independently but according to the probability of each letter, this time is slightly better, but still doesn’t capture the original structure. Then he uses second-order approximation, in this method, we pair each letter into two-letter pair, it has 3 states, each state begins with A, B, C., for example in the state B, we will find BA, BB, BC. Now that AA has more than one pairs, this is because AA in the source has a high probability of occurrence. Then we put each state into three cups choose each pair, and write down the first letter and choose the cup that defined by the second letter (Art of the Problem 2013).
Notice that this method makes it more similar to the original sequence. Then Shannon applies for third-approximation order, it is better, then he applies this to English words instead of a letter. Then he was thinking that how much information can we produce by using this process. To know which letter is going to produce next, instead think straightly, he cleverly converts this question to how many questions we are going to ask to know the next letter. For example, there are A, B, C, D, in order to know which letter is going to be, we need to ask 2 questions to know, if it's in the AB, if not, then ask if it's C or D. then he uses entropy to describe it as an uncertainty. To figure out the next letter, we need to ask 2 questions, so the information for A, B, C, D is 2 bits.

Technical quality

the technique methods Shannon uses are high quality, without doubt, he first uses the Markov model to elaborate his entropy theory, because the Markov model is a stochastic model used to model randomly changing systems (Markov model 2017). So, the future states will depend on current state rather than previous states, it ensures the model generates a random state. He uses this model to generate random letter, and later he finds out that by using different methods of zero, first, second-order approximation, he can improve the similarity to the original sequence of letters. This is closely tied up to his next finding, which is entropy.

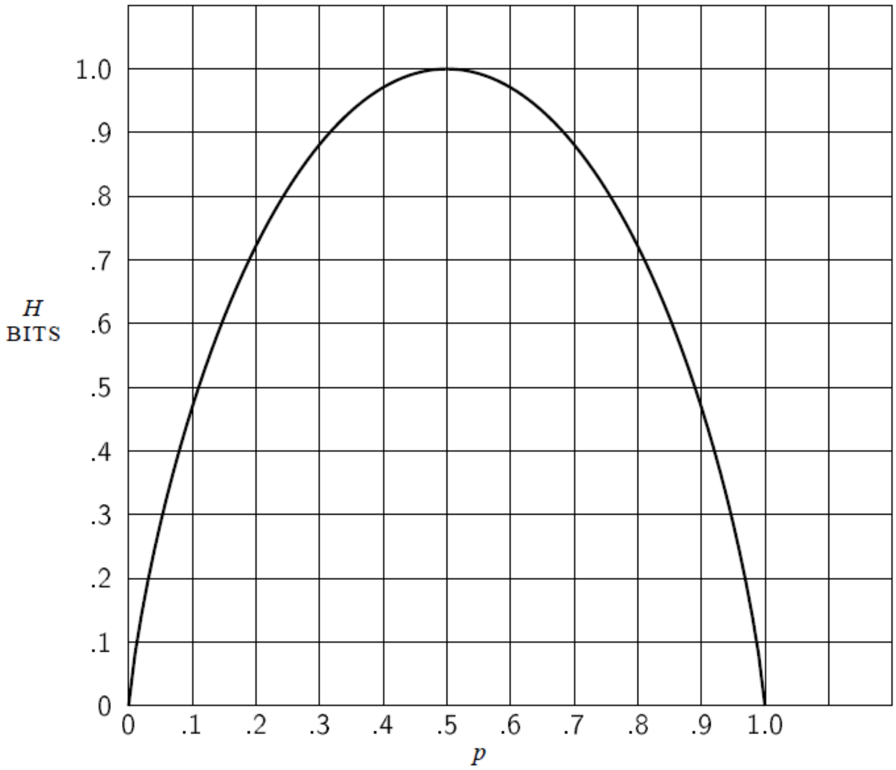

Fig. 7—Entropy in the case of two possibilities with probabilities p and $(1-p)$.

In figure 7, it explains how entropy change, when the probability is even, the entropy will be the biggest. All his theories have been proved two directions, which is necessary and sufficient condition. Shannon defines the two most important problems in communication and gives us the solution and prove that his solution is the best in mathematics (Art of the Problem 2013).

Application

The information theory can be applied to machine learning, computer science, machine communication because all these fields are involved with information. In other words, any fields that have to do with information can use information theory. in his paper, he talks about encoding formula, which helps when it comes to compression technique. Information is made up of code, there is so many redundant information it, by using an encoding technique, we can tremendously reduce the size of the information. Based on his work, it has developed Huffman Coding, it optimizes the compression technique further. for the channel capability part, it can be used in the telecommunication industry and internet communication. Once we get channel capability, we are not able to transfer information that exceeds the capability of the channel ($R <= C$), if we try to exceed it, it must lower our transmission rate. That could explain why the webpage can not be open even if we have already waited for a very long time. Shannon then says that it is always possible to find out an encoding method such that make the R infinitely close to C. but Shannon doesn't tell us how to find this encoding method. I think a way of finding out the most optimized encoding can be further researched in the future for his paper.

Presentation

The presentation style for this paper is coherent, in the beginning, Shannon introduces some background about communication, how communication history develops. he uses a lot of concepts to explain his discovery, such as Markov model, it's a little bit hard to follow easily, although it's a good example to illuminate his theory, still, it needs a lot of other knowledge to understand. the structure is well organized, each section talks about the specific problem. Each law could be strictly proved in mathematics. In general, it's a well-depth and groundbreaking paper.

Reference

Rabiner, L.R. 1989, 'A tutorial on hidden Markov models and selected applications in speech recognition', *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-86.

Art of the Problem, 2013, Information Theory part 11: Claude Shannon: A Mathematical Theory of Communication, viewed at 19 August 2019. < https://www.youtube.com/watch?v=3pRR8OK4UfE&t=1s>.

Art of the Problem, 2013, Information Theory part 12: Information Entropy (Claude Shannon's formula), viewed at 19 August 2019. < https://www.youtube.com/watch?v=R4OlXb9aTvQ>.

Markov model 2017, Wikipedia, viewed 18 August 2019, < https://en.wikipedia.org/wiki/Markov_model>.