College of Computer & Information Science
Northeastern University
CS G399: Algorithmic Power Tools I

Fall 2007
2 November 2007
Scribe: Emrah Bayraktaroglu

Lecture Outline:

- Introduction to Information Theory (Discrete Memoryless Channels)
- Zero-Error Capacity of a Communication Channel
- Pentagon

In this lecture, we will give an introduction to Shannon's mathematical model of information and a notion of the capacity of a channel. We will also analyze the zero-error capacity of the Pentagon graph, which is arguably the first problem that introduced semi-definite programming to computing.

1 Introduction to Information Theory

Let $\overline{x}$ be a vector of symbols that represents the input and $\overline{y}$ be the output vector we get when we send $\overline{x}$ over the channel C. It is highly possible that the channel would alter the input and cause $\overline{y}$ to be different that $\overline{x}$. That would lead to cases such as:

YOU WILL NOT ATTACK $\rightarrow$ YOU WILL NOW ATTACK

At this point, Shannon introduced DMCs (Discrete Memoryless Channels) to model the channels and information sent over them.

1.1 Discrete Memoryless Channel

In a discrete memoryless channel, we assume that messages are composed of symbols taken from a discrete domain and each message sent on the channel is independent of previous messages, thus making the channel memoryless. The information theory we will discuss in the class today is based on DMCs. We model information probabilistically, i.e., the amount of information according to this model is a function of the number of distinct messages that can be sent out. Formally, let $\overline{x}$ be a probability distribution over symbols and let P_x be the probability that $\overline{x} = x$. One can define *the surprisal* of symbol x as

$$\log\left(\frac{1}{P_x}\right)$$

. The *entropy*, or *uncertainty*, of $\overline{x}$, given by $H(\overline{x})$, is defined in terms of the surprisal of each symbol x as follows.

$$H(\overline{x}) = \sum_x P_x \log\left(\frac{1}{P_x}\right)$$

Intuitively, we are defining the amount of information in $\overline{x}$ in terms of the different values it can take; the *entropy* of a symbol space gives us the average amount of information we can get from a symbol.

Now it's time to define the information over the channel and capacity of a channel. We define it as the average amount of information we get about the input by observing the output. Let X be a random variable representing the symbol space for input and Y be the output random variable. We define the *mutual information* as

$$I(X,Y) = H(X) - H(X|Y)$$

Here $H(X)$ gives the uncertainty in the input and $H(X|Y)$ gives the uncertainty in the input, given the output. The difference between the two gives the information that has been provided.

And the capacity of a channel is the maximum information we can send over the channel. Capacity C is given as;

$$C = max_x I(X,Y) \tag{1}$$

Theorem 1. ***(Shannon's Noisy Channel Theorem)***

Every DMC's capacity C given by (1) is attainable asymptotically error-free.

By attainable, we mean that there exist encoding and decoding procedures such that capacity C is achieved and the resulting bit error rate can be made arbitrarily close to zero.

Around the same time that Shannon developed information theory, Hamming had also developed models for studying the communication of information. Instead of using a probabilistic model for errors, he considered bounds on number of errors in a message (e.g., at most d bits of a message are flipped), which perhaps made it more difficult to obtain general results.

2 Zero-Error Capacity

Shannon's Noisy Channel Theorem says that capacity C can be achieved with error that can made arbitrarily close to zero. It does not guarantee zero error however. Indeed, if every input symbol can be transformed by the channel to any other output symbol with non-zero probability, it is impossible to attain any zero error capacity. In his seminal paper, Shannon also introduced the concept of zero-error capacity.

Let $G = (V, E)$ be a graph, V set of vertices, which represent input symbols. Given $a, b \in V$, $(a, b) \in E$ if the symbols a and b cannot be confused with each other, i.e., the set of possible output symbols for a is disjoint from the set of possible output symbols for b. Any clique of G corresponds to a set of symbols that can be used simultaneously with zero-error. Thus, if we send one symbol at a time, we can attain a capacity of

$$\log_2 \omega(G)$$

with zero error.

But we can do better by considering blocks of symbols. Suppose we send messages out as blocks of n symbols. Then, we can attain a capacity of $\frac{\log_2 \omega(G^n)}{n}$, where the product $G^n = G \times G \times \ldots \times G$ (n times) is defined as follows.

$G' = G \times H = (V', E')$

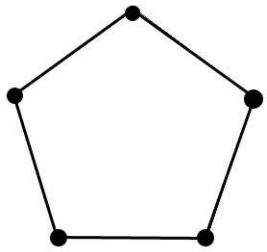

Figure 1: Representation of the pentagon (C_5)

$V' = \{(g_1 h_1) | g_1 \in V_g, h_1 \in V_h\}$

$E' = \{< (g_1 h_1), (g_2 h_2) > | < g_1, g_2 > \in E_g \vee < h_1, h_2 > \in E_h\}$

Zero-error capacity, also known as Shannon capacity is defined as:

$$SC(G^n) = \sup_n \frac{1}{n} \log_2 \omega(G^n).$$

It can be shown that $\omega(G^n)$ is supermultiplicative; that is, $\omega(G^{m \cdot n}) \geq \omega(G^m) \cdot \omega(G^n)$. This allows us to replace the supremum by a limit.

$$SC(G^n) = \lim_{n \to \infty} \frac{1}{n} \log_2 \omega(G^n) = \lim_{n \to \infty} \log_2 (\sqrt[n]{\omega(G^n)})$$

For convenience, the Shannon capacity is often defined as $\sqrt[n]{\omega(G^n)}$, which is what we will adopt in the remainder.

We do not know of any decidable procedure for computing the Shannon capacity of an arbitrary graph. In the following, we will try to obtain upper and lower bounds and calculate the capacity for a special graph, the pentagon. One simple upper bound on the Shannon capacity of G is $\log_2 |V(G)|$.

2.1 The Pentagon

This section will introduce the pentagon which will be used for capacity upper bound discussions in the next lecture. Pentagon is a graph C_5 that has five vertices each of which represent a symbol in the symbol domain connected in the shape of a pentagon. The figures 1 and 2 are representations for C_5 and (C_5^2), which yields (C_5) $\omega(C_5) = 2$ and $\omega(C_5^2) = 5$.

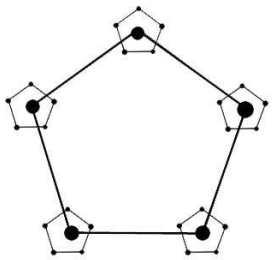

Figure 2: Representation for C_5^2