

Chapter 1

0 overview

1.1 Information Theory

We begin with the question, "What is a message and how is it communicated?" Messages are passed between people when they talk or write. The spoken form is probably the earliest, with writing coming much later. Today, messages are also passed between people and machines, between machines and people, and between machines.

Messages take many forms, such as acoustic pressure waves created by speaking or singing printed characters, drawings and paintings, and many more. How do we separate the message from the form in which it is presented? In what ways are a message and its form related? Marshall McLuhan¹ said the medium is the message. When is that true?

Information theory deals with both messages and their physical manifestations, which we call signals, from an engineering rather than a cultural viewpoint. The goal is to extend our capability to store and communicate messages by making efficient use of the storage and communication media. It also addresses the task of sensing the state of nature, such as detecting the presence of airplanes with a radar system, when the signals are weak and noisy.

We use a definition of a message that focuses on its symbolic and mathematical properties rather than its meaning. We intend that the most suitable

¹Marshall McLuhan (1911-1980) was a Canadian cultural critic and communications theorist who maintained that the method of communicating information has more influence on the public than the information itself. His books include *The Medium is the Message* (1967).

mathematical tools are those of probability and statistics, particularly of random processes. Messages can be thought of as arrangements of symbols from some alphabet. Statistical tools can be used to optimize message handling systems, because it is message traffic and not individual messages that are of interest for that purpose.

Messages must always appear in some physical form, which can be called a signal. A signal must have a unique correspondence to its message. The physical properties of the signal must match the medium that supports it. A good signal uses a small amount of the medium, is relatively immune to noise, and enables the message to be recovered easily and accurately.

Some signals occur naturally because they are generated by the message source as its way of making its messages observable. Speech is such a signal. Other signals arise because they are transformations that make it possible to carry the message through a different medium. Radio broadcast signals are an example.

The transformation between primary signals and secondary signals is of great interest, for it can endow the secondary signal with many special properties. In some cases, such as frequency modulation in radio broadcast, the goal is to make the signal more immune to noise by using more channel bandwidth. In other cases, such as speech compression or image compression, the goal is to produce a signal that is more compact than the original.

In many cases, particularly when digital processing is used, the primary signal may be transformed to a digital form. This is a secondary symbolic form with a very simple alphabet, namely, the binary numbers. This form can then be converted into a variety of physical signals by digital techniques. The binary representation has come to be the common form for essentially all forms of messages because of its convenient match with computing systems. Modern information theory therefore makes use of many techniques of digital signal processing, digital image processing and digital communications.

The common digital representation has made it possible to design digital communication devices, such as modems and digital telephone systems, and digital storage devices, such as magnetic and optical disks, that can be used with many kinds of signals. It is only necessary to convert a signal to digital form to be able to use these generic digital devices. This makes it possible to have a mass market for digital devices, which drives down unit cost. The dramatic drop in cost of magnetic storage devices is one example and the use of the CD-ROM, based on audio compact disks is another.

There is a common thread that runs through all of the systems, and that

is the foundation of information theory. By understanding the fundamental theory, it is possible to design systems that make the optimum use of the available resources and provide the best performance in particular applications. This theory is necessarily abstract, because it must be so if it is to be general. By being general, as well as useful, it has great power.

By focusing on the mathematical properties of messages rather than on their essence, it is possible to construct a set of powerful tools to guide the design of storage, communication and detection systems. We begin with our model for messages.

1.2 Message Model

Let us imagine the construction of a message by someone named Alpha who has something to communicate. Available to Alpha is a set of symbols that can be arranged in arbitrary sequences. Different arrangements of the symbols can represent different messages. This thinking leads us to the idea that a message can be represented by a particular arrangement of symbols from some alphabet.

To make a chosen symbol arrangement available to a friend, Beta, it is necessary that Alpha put them in a form that is observable by Beta. This form is constructed in the physical medium that is common to both Alpha and Beta. In the most basic case, each symbol can have a different mark in the medium. This is evident in writing where the symbols are the characters of the alphabet and the marks are shapes made with a marking instrument. These can be observed visually and interpreted symbolically.

If Alpha and Beta were to meet in circumstances such that they shared no common language or alphabet, it would be necessary to construct one. This could be done by a variety of means, and could be rapid and efficient if one of them had an understanding of information theory. The first step would be to construct a small set of basic symbols. The next step would be to relate short sequences of these symbols to physical objects and certain actions. These sequences correspond to words. By stringing together sequences of words it would be possible to convey a great many ideas. In a short time it would be possible to begin communication, and from there a culture would begin to emerge.

Note that the actual form of the symbols is not important. It is only the number of different symbols that is significant. If an alphabet has M letters,

then there can be a total of M^n different words of length n constructed with those symbols. Suppose that Alpha and Beta found a need for N different words. If all words were of the same length, the required length would be $n = \log N / \log M$. Any size alphabet with $M \geq 2$ would support their communication by making it possible to give each word a unique form.

The actual form of the symbols is not important. They must simply be distinguishable and easy to produce with the available instruments. We actually observe many different alphabets in use by different cultures around the world. We could use a similar approach to examine hieroglyphic systems, in which pictorial symbols are used to represent meaning or sounds or a combination of meaning and sound.

1.2.1 Binary Coding

Suppose that we have the task of converting messages to a binary form. Suppose that the messages are written in an alphabet with M letters. It would be a simple matter to give each letter a unique binary code and then convert the message letter by letter. This is exactly the approach used in ASCII². The approach is simple, but it is far from efficient. By taking advantage of the statistical properties of messages it is possible to find much more efficient coding schemes.

This leads us to the question of how efficiency can be measured or calculated. Suppose that one has two different binary coding methods called Code 1 and Code 2. If the codes are used to represent the same messages and Code 1 uses n_1 digits per 100 letters while Code 2 uses n_2 digits per 100 letters, we can find their relative efficiency by comparing n_1 and n_2 . However, to determine their absolute efficiency it is necessary to have an absolute standard. The absolute standard is provided by the entropy of the source. This is a statistical property that is fundamental to information theory. We will examine entropy in the next lecture.

1.2.2 Signal Properties

A signal is a physical property of a communication or storage medium upon which information is impressed by causing a systematic variation. In most

²American Standard Code for Information Interchange—a standard for defining codes for information exchange between equipment produced by different manufacturers.

communication channels the signal varies as a function of time, while in storage media it varies in spatial dimensions.

The variability of the medium can be characterized in terms of both the dependent and independent variables. These are related but separate characteristics. A signal $s(t)$ has a value s at each time t . The range of values of s that are available are determined by precision recording sensing and noise. If s is confined to a range $2A$ and can be resolved to a scale Δs , then a total of $m = 2A/\Delta s$ amplitude values are available at any particular time. This characterizes the variation that is available with the dependent variable.

The variation that is available with the independent variable, such as t , is determined by the rate at which the value of the signal may be changed. Suppose that the maximum rate of change of s is $a = \max(\dot{s})$. Then the signal would be able to change from one end of its range to the other in a time $\Delta t = 2A/a$. It would be possible, therefore, to record a new value of s about every Δt seconds. In a time of T it would be possible to record about $n = T/\Delta t = aT/2A$ values of s . Since there are $m = 2A/\Delta s$ choices at each Δt , the total number of recordings of length T are approximately $N(T) \approx m^n = (2A/\Delta s)^{(aT/2A)}$. This is of the form $N(T) = 2^{bT}$ where b is a constant.

The number of different binary patterns that can be supported are $\log_2 N = bT$. The rate at which the system can store or communicate data is $R = \frac{\log_2 N}{T} = b$. Some simple algebra shows that the binary rate that is available using this signaling scheme is

$$R = \frac{a}{2A} \log_2 \frac{2A}{\Delta s}$$

This is a constant that is determined by the properties of the medium and interface equipment. The ratio $a/2A$ is proportional to the bandwidth of the medium and interface equipment. We can therefore express the available rate as

$$R = kB \log_2 m$$

where B is the bandwidth and m is the number of distinguishable signal values.

In any communication or recording scheme, a major goal is to increase m and B . It is evident that increasing the bandwidth has much greater effect than increasing m . This is the reason it is common to speak of "bandwidth" when characterizing systems.

The above discussion has been in terms of the variation of a signal $s(t)$. A similar analysis can be done for signals that vary in spatial dimensions. Consider a signal $u(x, y)$. We can carry out an analysis that permits us to calculate the density of data storage per unit area. This computation will involve the same kind of considerations that we used above.

It is common to use a spatial system to record time-varying information. This can be done by relating x and y parametrically to t by prescribing a recording and playback path on the surface of the medium. At any instant the system is focused on a spot with coordinates $(x(t), y(t))$. From the viewpoint of the world outside the system, the signal looks like $s(t) = u(x(t), y(t))$. The path over the medium is completely hidden from the outside.

The above discussion is focused on the task of representing information for communication from Alpha to Beta. It does not address questions such as the meaning or value of the information. So, if we do not address anything related to the content of messages, what is the value of information theory? The answer is that it enables us to understand how to construct efficient and reliable systems.

In addition to the techniques for the transmission of messages, information theory provides us with a means to describe information sources. This is very important in characterizing such natural sources as images and speech and text. A good model of a natural source is necessary to be able to represent its message efficiently, and is useful in systems for content understanding.

1.3 Source Modeling

The simplest model of a source is a system that selects one message from a certain set. If the set of messages is of size M , then it is necessary to have M distinguishable signals that can represent them. We will find that even in this simplest of all cases, the calculation of the best code is not obvious. David Huffman won fame for solving this problem.

Of course, it does not seem sensible to construct an efficient code for a source that only produces one message in its life. We are usually considering sources that generate an ongoing sequence of messages. In that case, we are interested in questions such as the average length of the signal for each message, the average number of bits per second, the probability of error, and so forth.

The next most sophisticated model is one in which message context is

important. An example is a device such as a teletype that is printing a message a character or a word at a time. The message you have seen so far gives you information about what might come next. Modeling this behavior requires accounting for the linkages within the messages. The simplest model of this type is probably the Markov model. Markov models have been used for English text as well as for human speech. The heart of modern speech recognition systems is the hidden Markov model (HMM), a very sophisticated Markov source model.

Images are a type of information for which efficient coding is very important because of the sheer size of image files. The quest for good codes to represent images falls under the heading of image compression. By compression we mean finding a better code to represent an image than would be achieved by a simple direct coding of some sort. Image compression can be lossless or lossy, depending on whether the original can be obtained exactly or approximate from the code. An example of a lossless code is a Huffman code or a LZW code. The format JPEG is an example of a lossy code.

Image sequences are even more demanding of communication bandwidth and digital storage. Progress in the coding of image sequences has made digital television possible. There has been a steady advance in techniques to encode video images and sound. A very sophisticated standard called MPEG-7 is about to be implemented. This will open a broad range of applications.

It is not unreasonable to assert that none of the advances described above would have been possible without the conceptual tools that are made available by information theory. The importance is about equally divided between source coding and channel coding. We shall strive to develop a good sense of both in this course.

