Information Theory and Coding
Principles of Information theory

م. فؤاد حمادي

2020 - 2019

Principles of Information theory

• Introduction

Data compression addresses the problem of reducing the amount of data required to represent a digital file, so that it can be stored or transmitted so efficiently.

The principle of data compression is that, it compress data by removing redundancy from the original data in the source file.

On the other hand, information theory tells us that the amount of information conveyed by an event relates to its probability of occurrence. An event that is less likely to occur is said to contain more information than an event that is more likely to occur.

The problem of representing the source alphabet symbols (usually the binary system consisting of the two symbols 0 & 1) is the main topic of **coding theory**.

An optimum coding scheme will use more bits for the symbols that less likely to occur, and a fewer bits for the symbols that frequently occur. coding theory leads to information theory and ***information theory*** provides the performance limits on what can be done by suitable encoding of the information. Thus the two theories are intimately related. Both ***coding and information*** theory give a central role to errors ***(noise)*** and are therefore of special interest since in real-life noise is everywhere.

The conventional system is modeled by:

1- An information source.

2- An encoding of this source.

3- A channel over or through which the information is sent.

4- A noise (error) source that is added to the signal in the channel.

5- A decoding and recovery of the original information from the received signal with noise.

6- A destination for the information.

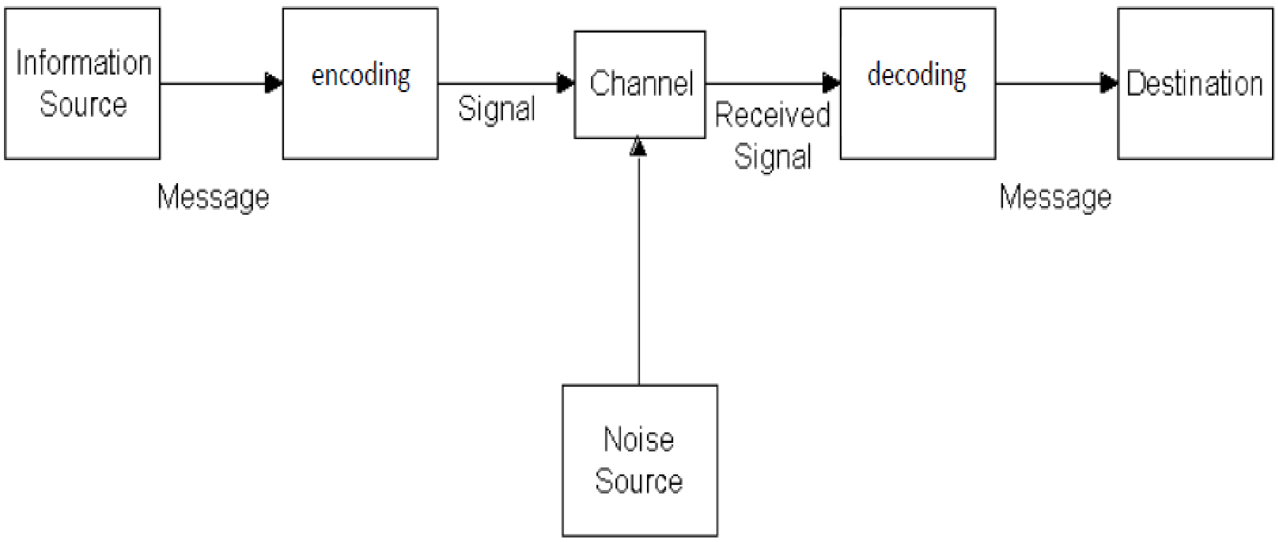

FIGURE 1 communication system

All communication systems involve three steps:

1. Coding a message at its source

2. Transmitting the message through a communications channel

3. Decoding the message at its destination.

- In the first step, the message has to be put into some kind of ***symbolic representation*** – words, musical notes, icons, mathematical equations, or bits.
- ***Transmission*** can be by voice, a letter, a billboard, a telephone conversation, a radio or television broadcast.
- At the destination, someone or something has to receive the symbols, and then decode them by matching them against his or her own body of information to extract the data.

• Information Theory

Information theory tells us that the amount of information conveyed by an event relates to its probability of occurrence. An event that is less likely to occur is said to contain more information than an event that is more likely to occur. The amount of information of an event and its probability are thus opposite.

Information Theory is a mathematical subject dealing with ***three basic concepts:***

1- The measure of information.

2- The capacity of communication channel to transfer information.

3- Coding as a means of utilizing channels at full capacity.

These three concepts are tied together in what can be called the fundamental theorem of information theory as follows: Given an information source and a communication channel, there exists a coding technique such that the information can be transmitted over the channel at any rate less than the channel capacity and with arbitrarily frequency of error despite the presence of noise.

- ## Information Source

Information theory uses the entropy function H as a measure of information, and by implication this defines that is meant by "the amount of information". Information theory ***does not handle the meaning*** of the information; it ***treats only the amount*** of information.

The source of information may be many things. For example: a ***book***, a ***printed formal notice*** and a ***financial report*** are all information sources in the conventional alphabetic form. Information also exists in ***continuous forms***, but modern practice is to ***digitize*** the continue forms by ***sample*** the continuous signal at equally spaced intervals of time and then to ***quantize*** the amount observed. The information is then sent as a stream of digits. The ***first reason*** for this use of digital sample of the analog signal is that they can be transmitted ***more reliably*** than can the analog signal. When the inevitable noise of the transmission system begins to degrade the signal, the digital pulses can be sensed (detected), reshaped and amplified to standard form before relaying them down the system to their final destination. At the destination the digital pulses may,

if necessary, be converted back to analog form. A ***second reason*** that modern system use digital methods is that ***integrated circuits*** are now ***very cheap*** and provide a powerful method for flexibly and reliably processing and transforming digital signals.

- ## Encoding a Source Alphabet

It is conventional to represent information as being in one of two possible symbols "0 " an d "1 " . At present devices with two states, called binary devices. For representing the various symbols of the source alphabet:

1.Use two binary digits, we can represent four distinct states (symbols): 00 01 10 11

2. For three binary digits we get $2*2*2=2_3$ =8 distinct states (symbols) 001 010 100 110 001 011 101 111

3. For a system having **k** binary digits – usually abbreviated as bits-- the total number of distinct states is $\mathbf{2_k}$ In general, if we have **k** different digits, the first having $\mathbf{n}_1$ states, the second $\mathbf{n_2}$ states, … , the $\mathbf{k_{th}}$ having $\mathbf{n_k}$ states, then the total number of states is clearly the product $\mathbf{n_1 * n_2 * n_3 * \ldots * n_k}$

Example: let $\mathbf{n_1}$ represent an octal digit, n_2 represent a decimal digit, and $\mathbf{n_3}$ represent a hexadecimal digit, then we can represent

8 * 10 * 16 = 1280 distinct state (symbol).

This is the number of source symbols we can represent if all we consider is the number of distinct states.

- **Some Particular Code**

Binary code is difficult for human being to use. People prefer to make a discrimination among many things. Such as decimal system with 10 distinct symbols. Thus for human use it is often convenient to group the binary digits, called bits into groups of three at a time and call them the octal code (base 8). This code is given in the table below:

Binary	Octal	Binary	Octal
000	0	100	4
001	1	101	5
010	2	110	6
011	3	111	7

For example, the decimal number 25 is written into octal as: $(25)_{10} = (31)_8$ The translation form octal to binary is so immediate that there is a little trouble in going either way:

$(31)_8 = 011\ 001$ which is equal to $(25)_{10}$

The binary digits are grouped in four to make the hexadecimal code, as

Binary	Hexa	Binary	Hexa	Binary	Hexa
0000	0	0101	5	1010	A
0001	1	0110	6	1011	B
0010	2	0111	7	1100	C
0011	3	1000	8	1101	D
0100	4	1001	9	1110	E
				1111	F

- ## Some of the International Coding Techniques

1. The ASCII Code (American Standard Code for Information Interchange) :

Given an information source, we first consider an encoding of it. The standard ASCII code (**American Standard Code for Information Interchange)** and it is most widely used with microcomputers, table below, which represents alphabetic, numeric and other symbols in an example of a code. Basically, this code uses seven binary digits. Since computers work in bytes which are usually block of 8 bits, a single ASCII symbol often use 8 bits. The eighth bit is called the parity check bit , It is usually set so that the total number of 1's in the eight positions is an even number (or else an odd number).

Char ASCII	Code	Char ASCII	Code	Char ASCII	Code
A	**(41)**$_H$	J	**(4A)**$_H$	S	**(53)**$_H$
B	**(42)**$_H$	K	**(4B)**$_H$	T	**(54)**$_H$
C	**(43)**$_H$	L	**(4C)**$_H$	U	**(55)**$_H$
D	**(44)**$_H$	M	**(4D)**$_H$	V	**(56)**$_H$
E	**(45)**$_H$	N	**(4E)**$_H$	W	**(57)**$_H$
F	**(46)**$_H$	O	**(4F)**$_H$	X	**(58)**$_H$
G	**(47)**$_H$	P	**(50)**$_H$	Y	**(59)**$_H$
H	**(48)**$_H$	Q	**(51)**$_H$	Z	**(5A)**$_H$
I	**(49)**$_H$	R	**(52)**$_H$		

Seven-bit ASCII Code

The ASCII code is a fixed length code where all the code words are of the same size.

2. Morse Code:

Morse code is one of the familiar codes, which was once widely used. As noted earlier most systems of representing information in computers (and other machines) use two states, although the Morse code is an example of signaling system with three symbols in the underlying alphabet. The Morse code is clearly a ***variable-length*** code which takes advantage of the high frequency of occurrence of some letters such as “E”, by making them short and the very infrequent letters such as “J”, relatively longer.

Symbol	Morse Code	Symbol	Morse Code
A	.-	N	-.
B	-...	O	---
C	-.-.	P	.--.
D	-..	Q	--.-
E	.	R	.-.
F	..-.	S	...
G	--.	T	-
H	....	U	..-
I	..	V	...-
J	.---	W	.--
K	-.-	X	-..-
L	.-..	Y	-.--
M	--	Z	--..