INFORMATION THEORY AND COMMUNICATION

Tibor Nemetz
Rényi Mathematical Institute, Hungarian Academy of Sciences, Budapest, Hungary

Keywords: Shannon theory**,** alphabet, capacity, (transmission) channel, channel coding, cryptology, data processing, data compression, digital signature, discrete memoryless source, entropy, error-correcting codes, error-detecting codes, fidelity criterion, hashing, Internet, measures of information, memoryless channel, modulation, multi-user communication, multi-port systems, networks, noise, public key cryptography, quantization, rate-distortion theory, redundancy, reliable communication, (information) source, source coding, (information) transmission, white Gaussian noise.

Contents

Summary

Information theory is a mathematical theory that quantifies information and utilizes these quantities for modeling situations and solving optimality problems of communication and information storage. It deals with both theoretical and practical aspects of data compression and reliable transmission of information over noisy channels. The data source entropy gives a lower bound for the rate of data compression. Rates for reliable information transmission are bounded by the capacity of the given channel. The theory also includes theoretical analysis of secrecy systems.

1. Introduction

1.1. The Origins

The practice of *telecommunication*, i.e. transmitting messages at a distance without human messenger, has very old roots. The roman historian Polybios (210-128 B.C.) describes a wireless *telecommunication system*, which was applied for transmitting texts written in the 25 letters Latin alphabet. This system could be termed as torchlight broadcasting. We use it to describe some basic notions.

The *transmitting station* consisted of two walls of 5 holes in each. The two walls were built at a distance allowing clear differentiation between them. Sending a *source-message* was performed by *signaling* a combination of torch lights in each wall according to the subsequent letters of the text. According to the rules, only the number of torches was important, not their physical place. This allows us to represent the signals by pairs of the digits {1, 2, 3, 4, 5}. We may term this set of 5 digits as *code alphabet*. For each letter of the *source alphabet*, it was necessary to fix the pair of code letters (digits) which represented the given letter. This was described by a 5x5-table code. This table could be considered as the early appearance of *block codes*.

	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	I	K
3	L	M	N	O	P
4	Q	R	S	T	U
5	V	W	X	Y	Z

Table 1: Polybios' alphabet

The *encoding* was performed letter-by-letter. Then the *encoded text* was transmitted with an agreement that the rows correspond always to one of the two walls, the columns to the other. The *receiver* could observe these signals, and knowing the role of the walls, he could *uniquely decode the received message.*

Of course, the signaling could be observed by anybody, not only the intended receiver. Therefore there was an obvious need to hide the information contained in the message. To this end a secret transformation was applied to the clear text. The emperor Caesar (100-44 B.C.) applied a very simple secret *encryption* code: He has substituted the letters by the cyclically third letter of the alphabet, e.g. A by D, Z by C. Such simple substitutions are *called Caesar substitution*s even if the letter replacing a clear text letter is the subsequent *k*-th and not the third.

Several similar *ad hoc* codes were designed and applied. Historically the next one to be mentioned is the Morse telegraph (1830). The painter Samuel Morse (1791-1872) invented an electro-magnetic way of sending written messages. The sender and the receiver are connected by a wire. The sender initiates short or long electronic impulses and the receiver perceives these signals. The wire is an example for a *transmission channel* with identical input-output alphabet. This alphabet consists of 3 symbols: a short impulse (usually represented by a point), followed by a "lack of impulse" of the same duration, a long impulse (represented by a dash) of 3 times longer duration, followed by the "lack of impulse", and a 4 times longer "pause (comma)". Letters, digits and other punctuation marks are represented by different *variable length* sequences of points and dashes. The correspondence is called *Morse code*. The pause is used to separate these sequences; therefore any code-sequence can be uniquely decoded. This property would not be valid without applying the separating commas. The code shows Morse's intention to minimize the time of transmission. The sequences representing more frequent letters need shorter time, the less frequent ones longer time. This economical consideration is made mathematically precious within the theory.
Soon after Morse's invention, a lawyer published a *codebook* allowing more effective *compression* of special texts. He has collected typical fragments of commercial correspondence. These fragments were assigned 3 symbol blocks, and these blocks replaced the fragments in letters. The idea was not new: military applications of codebooks are known already from the 17th century.

Further milestones of the telecommunication practice are

- Telephone: In 1861 Phillip Reis German physicist developed an instrument for transmitting sound/voice/music as electronic waves. He has named his instrument *Telephone*. In 1876 Alexander Graham Bell (1847-1922) has perfected it. The essence of the discovery is that sound-waves can be transformed into electronic waves, which could be forwarded on electric wire.

- Guglielmo M. Marconi (1874-1937), summarizing the results of several experiments, has invented the *Wireless Telegraph* in 1901. This contained already the basics for the radio. The area of modulation was born.

- The first *commercial radios* started their regular program in the USA and Germany in 1920.

- The ideas of broadcasting pictures (beside voice) became a reality around 1925. The first TV broadcasting was in London, early 1926 (by John L. Baird). For a functioning solution, problems of synchronization had to be solved.

- The idea of combining telegraph and printing machines lead to the development of *teletype communication* around 1930. This system applies a *binary block code* of length 5. A binary alphabet consists of two letters, usually represented by 0 and 1. The expression binary digit is shortened into *bit*, a basic notion of computer and communication practice and theory. By the 5 bits 32 combinations can be formed; however there are more symbols on the keyboard of a typing machine. Therefore the encoding is performed as a two-state machine. There is one-one transformation of 29 symbols into 29 binary combinations in both states, while 2 of the 32 combinations are used to indicate the state. This is called telex-code and is fixed by international standard.

- The idea of representing analog symbols as digital sequences appeared in the second half of the 1930s. The first *Vocoder* was constructed in the Bell Laboratories in 1939.

Systematic analysis of the area of telecommunication started during World War II. The work started by recognizing that the performance of communication systems is highly influenced by noise. A formal description of both the noise and the signals was needed. The solution was that these were modeled as random processes. The Russian V. A. Kotyelnikov made research on detection and estimation of signals at the receiver (published in 1947). N. Wiener worked on linear filters separating the noise from additive noise (published in 1949). More systematical investigation was carried out by C. A. Shannon in the early 1940's. He has published "A Mathematical Theory of Communication" in 1948. The publication date is referred to as the birthday of modern information theory. In another paper he has summarized the existing knowledge, building a complete "Communication theory of secrecy systems" (1949).

1.2. Shannon Theory

Information theory was created by Claude E. Shannon for the study of certain quantitative aspects of information, mainly as an analysis of the impact of coding on information transmission. His key observation was that the colloquial term "information" needs no exact definition. Semantic aspects are irrelevant for communication engineering. Rather it needs operational characterization through numerically measurable quantities.

Another important viewpoint was that information sequences are random processes; therefore probability theory should be applied when analyzing problems concerning information systems. He named the emerging theory "Information theory", a name, which has become commonly used in the communication society. It comprises also all subsequent theoretical and practical results. Theoretical results in the narrower areas initiated by Shannon are usually referred to as the Shannon theory.

He has introduced the notion of entropy as a measure of uncertainty (or information). Entropy as a measure of the amount of information allows the formulation of the solution of many important problems of information storage and transmission. The operational interpretation of the entropy is very concrete: Roughly, it equals the minimum number of binary digits needed, on the average, to encode in a uniquely decodable way a given message.

Although the introduction of entropy was motivated mainly by coding problems, several natural postulates lead to the same quantity, as was firstly shown by D. K. Fadeev in 1956.

1.4. Future in Present

Present development of technology is too fast for the theory to follow, although the Shannon theory provides suitable orientation. Plenty applications are ad-hoc and their deeper analysis could certainly lead to a new theory. Electronic correspondence (e-mail), Internet connections apply several protocols, which are not present in the existing theories, but their use is a common practice now. The increasing communication speed calls also for theoretical consideration. Further a huge area is quantum communication and quantum information theory. ISDN packet communication may also lead to new approaches. New forms of Mobile telephone (SMS) need also some theoretical analysis.

2. Information Source

Information sources are identified in Shannon's theory by their outputs. It is supposed that an information source outputs a sequence of finite random variables over a finite set, called *source alphabet*. The source is known if all the finite dimensional distributions of the random outputs are known. The most important distributions correspond to the independent, identically distributed random variables. Such sources are called *discrete memoryless sources, abbreviated as DMS*. The first theoretical results explained by Shannon concerned this case. The results were extended to the *Markovian* case, later to the more general *stationary ergodic processes*.

The source statistics need not be known exactly. The theory can be applied under the hypothesis that it is known, while the probabilities are replaced by their statistical estimation. The theory covers the cases, where different source sequences follow different statistics. Such an example is the letters in a book from an international library. The statistical attitude, neglecting semantic aspects, is heavily used in constructing models, in formulating optimality criteria and in solving them.

3. Source Coding

For many possible reasons the source output sequence needs to be converted into a sequence over another finite set, called code-alphabet. The way of converting is called *encoding*. One may wish to reconstruct the original source output from the encoded version. The rule for reconstructing is called *decoding*. The source- and the code alphabets may be identical, which is the typical case of secrecy systems. They may be different like in the case of the Morse code. In data storage applications the code-alphabet

contains two elements, the binary digits 0 and 1, called bits only. Such codes are called *binary codes*. We will discuss binary codes, but the results can be extended to any finite code-alphabet size.

3.1. Uniquely Decodable Codes

Codes, which allow *perfect decoding*, i.e. when all encoded sequences can be reconstructed without error, are called uniquely decodable codes. The simplest case is the case of the *letter code*. This encodes all source outputs into a binary string, called *code word*, or simple word. The source sequence is then encoded letter-by-letter by concatenating the corresponding code words. The code is called *block-code* if all the code words are of the same length. Code 1 on Table 2 is a block code. The block-codes are obviously uniquely decodable. A code is a *variable-length code*, if the lengths of the code words are not the same. Codes 2, 3 and 4 are of variable length. Codes 2 and 3 are uniquely decodable but Code 4 is not.

Source letters	Code 1	Code 2	Code 3	Code 4
A(0.25)	000	00	00	00
B(0.25)	001	01	001	01
C(0.10)	010	100	101	000
D(0.10)	011	101	1011	001
E(0.10)	100	110	1111	100
F(0.10)	101	1110	0011	101
G(0.05)	110	11110	00111	110
H(0.05)	111	11111	10111	111

Table 2: Examples of letter codes

If a sequence is encoded by a variable length binary code, then it starts with one of the code words. Therefore the decoding may start with searching for this word. If it is unique, then the corresponding source letter was encoded, by necessity. There can not be two different words at the beginnings, if no code word is the beginning of any other code words. In such cases the decoding can be continued from the end of the first code word, and so on. Codes with this property are called *prefix codes*. Code 2 is a prefix code. The prefix property is a sufficient condition for unique decodability, but not necessary. Code 3 is not prefix, since the word of the source letter A is the beginning of the word of B. Nevertheless, this is also uniquely decodable: Looking at the code-words in a reversed order (from backwards) the binary sequences form a prefix set. Therefore if we start the decoding from the end and work towards the beginning, we arrive at a single letter sequence, the one, which was encoded. Encoding the source output AE by Code 4 one gets the binary string 00-100. This string can be parsed into 001-00, and this shows that the source sequence DA yields the same binary string, therefore Code 4 is not uniquely decodable.

L.G. Kraft has characterized the binary prefixed codes. Suppose, that the source alphabet contains K letters, and there is a binary prefix code with word-lengths $N_1, N_2, \ldots, N_K$. Then the inequality

$$\frac{1}{2}N_1 + \frac{1}{2}N_2 + \ldots + \frac{1}{2}N_K \leq 1 \qquad (1)$$

holds true. On the other hands, if the numbers $N_1, N_2, \ldots, N_K$ satisfy this inequality, then there is a binary prefix code with such word-lengths. This assertion is referred to as *Kraft inequality*. Later, in 1956 B. McMillan proved that the same inequality holds for any uniquely decodable code. In this case the inequality is referred as *McMillan's inequality*.

Beside the demand that a code should be uniquely decodable, another important aspect should be considered when choosing a specific code. The number of bits in the encoded sequence should be as small as possible. With block codes, this number is proportional to the block-length. In the case of variable length codes, this number is a random variable. In such cases the expected word-length should be minimal. Since the expected value is a linear functional, it suffices to consider letter-codes, only. A code is better than another one, if its expected word-length is smaller than that of the other. In the examples of Table 2 the first column specifies source output probabilities. Code 1 is a block code, the word-length is 3. The expected word-length of Code 2 is 2*0.5+3*0.3+4*0.1+5*0.1=2.8 < 3, therefore Code 2 is better than the block-code 1.

One of the basic questions of information theory is finding the "best" code, i.e. the code for which the expected word-length is minimal. The solution was given by D. A. Huffman in 1952. His main idea was that there is an optimal code with the property that the two letters of the smallest probabilities have equal word-length. These two letters may be combined into a single super-letter. The super-letter's probability is the sum of the two smallest probabilities. Then one has to find an optimal code for this modified source, and suffix the code word of the super-letter by a 0 and by a 1 to get code words for the original letters. This algorithm reduces the alphabet size step-by-step, till there remains a two-letter alphabet. This is a trivial case, where the best code is a "1-bit block code" Going through the steps in reversed order, we construct *one* optimal code. The algorithm is called *Huffman algorithm*, the resulting codes are called *Huffman codes*.

Shannon, in his basic work defined another variable length code. He has arranged the alphabet according to descending order of their probabilities. Then the alphabet was spilt into two subsets in a way that both subsets had probabilities as near to one half as possible. Letters in one subset were assigned codes starting with 0, the code words of letters in the other started by 1. Each subset was then recursively further divided till all subsets had just one element. This method was also discovered independently by Fano, and is now referred as *Shannon-Fano code.*

The entropy may be approached by using long block codes. For all block-lengths, one should find the Huffman code. Arithmetic codes spare this difficulty. As an extreme, here the block-length may be as long as the entire source output, without caring about the definition of the code. Arithmetic coding for discrete memoryless sources goes letter-by-letter, recursively. It orders subintervals of the unit interval to source sequences. It starts by dividing the [0,1) interval according to the source probabilities, and then chooses the subinterval corresponding to the actual source output. Then the process is repeated with the subinterval chosen, dividing it and choosing one sub-sub interval according to the new source output. The result is a subinterval [L,R) with left-end point L

and right-end point R. Now, consider the binary representation of an arbitrary point of this interval. Take as many bits of this representation, which yield a number within the interval. The smallest bit string will be used as the arithmetic code for the given source sequence. The number represented by this binary string is an inner point of all intermediate subintervals. Therefore the decoder may repeat the procedure and find the subsequent source output one after the other. Arithmetic coding was developed by J. Rissanen and R. Pasco in 1976.

-
-
-

Bibliography

Blahut R. E. (1983). *Theory and Practice of Error Control Codes*, Addison-Wesley, Reading, MA. [This book has been chosen from a number of good books on error detecting and error correcting codes as theoretically sound, and at the same time easy-to-follow information source.]

Csiszár I., Körner J. (1981*). Information Theory. Coding Theorems for Discrete Memoryless Systems*. Budapest: Akadémiai Kiadó, New York: Academic Press, 452 pp. [The book presents a well-integrated mathematical discipline, concentrating on quantitative aspects of coding for information transmission and storage in the case of discrete memoryless systems.]

Hamming R. W. (1950). Error Detecting and Error Correcting Codes. *Bell System Techn. Journal*. Vol. 29, 47-160. [This is the first systematic treatment of error control problems.]

Kahn D. (1967): The codebreakers: The story of Secret Writings. New York: McMillan. Abridged ed., (1973): New York: New American Library. [This "bible" of cryptography follows all important moments of the history and practice of cryptography. Written by a journalist, it provides an enjoyable reading.]

Kullback S. (1959). *Information theory and statistics*, New York: John Wiley & Sons. [This book offers a detailed mathematical analysis of application of information theoretical methods to statistical decision problems.]

Salomon D. (1998). *Data Compression: the Complete Reference*. New York: Springer. [This book provides an overview of the many different types of compression. It includes taxonomy, an analysis of the most common methods of compression, discussion of their benefits and disadvantages, and their most common usage. Detailed descriptions of the most well-known compression methods are covered.]

Shannon C. E. (1948). A mathematical model of communication. *Bell System Technical Journal*, Vol. 27, pp. 379-423 and 623-656. [The first systematic mathematical analysis of communication practice leading to an abstract mathematical model. Its publication date is considered to be the birthday of information theory.]

Shannon C. E. (1949). Communication Theory of Secrecy Systems. *Bell System Technical Journal*, Vol. 28, pp. 656-715. [This paper establishes the sound theoretical basis for analyzing cryptographic systems, analysing and abstracting past experiences. It lays down the fundamentals for modern cryptology.]

Simmons G. J. (ed). (1991). Contemporary Cryptography. New York: IEEE Press. 640 pp. [This collection of 13 papers provides an easy-to-read up-to-date summary of the essential state of art in the main areas of cryptology: Cryptography in general, Authentication, Protocols, Cryptanalysis, and Smart card area.]

Verdú S. (ed.) (1998). Information theory: 1948-1998. Special Commemorative Issue. *IEEE Transactions on Info. Theory*, vol. 44, No.6, 2042-2772. [25 invited papers summarize the fifty years of developments of the Shannon theory and/or its important sub-areas.]

Biographical Sketch

Dr. Tibor Nemetz, DSc., has been working in the field of information theory and cryptography since 1964. He is a full professor at the Eötvös Lorand University, Budapest, and senior scientific adviser at the Rényi Mathematical Institute of the Hungarian Academy of Sciences. He was visiting professor at the Carleton University, Ottawa, Canada, J.W. Goethe University, Frankfurt, Germany, Technische Universität of Wien, Austria, and at the Middle East Technical University, Ankara, Turkey. His university courses include Information Theory, Mathematical Statistics, Probability Theory, and Data Compression. In 1989 he has offered the first officially permitted university credit course in crytography in Hungary, and published the first Hungarian scientific book on *Algoritmic Datasecurity*, co-authored by I. Vajda. He was chief organizer of Eurocrypt '92 of the IACR. He was invited to deliver talks at a number of international and national conferences. He has published more than 100 scientific publications.