THE ANNALS OF "DUNAREA DE JOS" UNIVERSITY OF GALATI
FASCICLE III, 2006 ISSN 1221-454X
ELECTROTECHNICS, ELECTRONICS, AUTOMATIC CONTROL, INFORMATICS

OPTIMAL CODING FOR QUALITATIVE SOURCES ON NOISELESS CHANNELS

Valeriu Munteanu and Daniela Tarniceriu

Faculty of Electronics and Telecommunications, Bd. Carol I, no. 11, Iasi, Romania

Abstract: In this paper we perform the encoding for sources which are only qualitatively characterized, that is, each message the source delivers possesses a certain quality, expressed as cost, importance or utility. The proposed encoding procedure is an optimal one, because it leads to maximum information per code word and it assures a minimum time for the transmission of the source information.

Keywords: coding, information theory, transmission channels.

1. INTRODUCTION

In the classical source encoding procedure on noiseless channels (Huffman, 1953; Gallager, 1968; Cover, and Thomas, 1991), each message m_j is assumed to be delivered with the probability p_j. In these encoding procedures the quality (cost, importance or utility) of the messages the source delivers, are not considered. In this case the optimal codes assure a minimum average length of the code words or a minimum rate. In many real cases the messages possess some qualities for the receiver. In these situations the message probabilities are not specified. Each message m_j is possessed of a real number u_j, that measures either the cost, or the importance or the utility of that message from the receiver point of view. For instance, we want to transmit the code words $c_i, i=1,2,...,n$, of lengths l_i. Obviously, the larger the codeword length is, the larger the transmission time is, and, therefore, the cost to transmit it will be larger. Similarly, if we want to transmit a price cutting for market articles, obviously the buyers will consider more important the ones which have significant cost reduction. Two problems arise while transmitting messages which contain costs or utilities. The first one is to reach the maximum information per code word, and the second one is to minimize the transmitting time for the source information.

To solve these two requirements, in Section 2, we will consider that the messages m_j, $j=1,2,...,n$, possessed of the costs or utilities u_j and the probabilities p_j are determined, so that the information per code word is maximum. To this aim, we will derive the mean quantitative – qualitative information per message for a source which delivers the messages m_j, $j=1,2,...,n$, possessed of the utilities u_j. Attempts to find out quantitative – qualitative entropies are given in (Belis, and Guiasu, 1968; Khan, and Autar, 1979; Autar, and Khan, 1989; Gurdial, and Pessoa, 1977; Munteanu. and Cotae, 1992; Longo, 1976; Khan et al., 2005).

In this paper we infer a quantitative – qualitative entropy that differs from the ones known until now. The main benefits of this entropy consist in:

1- If the qualitative characteristic of information is neglected, the entropy we defined coincides with the entropy defined by Shannon (Shannon, 1948).

This paper was recommended for publication by Dorel Aiordachioaie

2- If the utilities or costs are assumed to be real numbers, the main properties of the classical entropy are preserved by the quantitative – qualitative entropy defined in this paper.

The entropies given in (Belis, and Guiasu, 1968; Khan, and Autar, 1979; Autar, and Khan, 1989; Gurdial, and Pessoa, 1977; Munteanu. and Cotae, 1992; Longo, 1976; Khan et al., 2005) do not satisfy these properties.

Once the optimum probabilities with which the source messages have to be transmitted, to assure the minimum length per code word are determined, any classic encoding procedure can be used (Huffman, 1953; Gallager, 1968; Cover, and Thomas, 1991). In Section 3 a binary Huffman encoding is performed for a source characterized only qualitatively and the average length of the code words is computed. Section 4 summarizes briefly the main conclusions of this paper.

2. DETERMINING THE QUANTITATIVE – QUALITATIVE ENTROPY

Let S be a discrete, memoryless source characterized by the distribution:

$$(1)\ S:\begin{pmatrix} m_1 & m_2 & \cdots & m_n \\ u_1 & u_2 & \cdots & u_n \end{pmatrix}$$

where u_i measures the quality (cost, importance or utility) of the message m_j .

The question that arises is how to transmit the information of this source on a noiseless channel, so that the information per message is maximum one and the corresponding transmitting time, minimum. To answer to these two requirements we assume that the messages $m_j,\ j=1,2,...,n$ are transmitted on the channel with the probabilities p_j. To determine them, we have to find out the quantitative – qualitative entropy of the source characterized by the distribution

$$(2)\ S:\begin{pmatrix} m_1 & m_2 & \cdots & m_n \\ p_1 & p_2 & \cdots & p_n \\ u_1 & u_2 & \cdots & u_n \end{pmatrix}$$

and then to find out the probabilities p_j, which maximize it.

Theorem 1- The quantitative – qualitative entropy of the source characterized by (2) is given by

$$(3)\ H_{pu}(S) = -\sum_{j=1}^{n} p_j \log_2 p_j + \sum_{j=1}^{n} p_j u_j$$

Proof: Let σ_k be the event consisting in the delivery of the messages m_i and m_j, that is

$$(4)\ \sigma_k = (m_i m_j)\ ,$$

Assuming the source is memoryless, that is, the messages m_i and m_j are probabilistic and logic – causal independent, we have

$$(5)\ p(\sigma_k) = p_i p_j$$

$$(6)\ u(\sigma_k) = u_i + u_j$$

The quantitative – qualitative information attached to σ_k, denoted by $i_{pu}(\sigma_k)$, is, generally, a function of the probability $p(\sigma_k)$ and the utility $u(\sigma_k)$, i. e.

$$(7)\ i_{pu}(\sigma_k) = F[p(\sigma_k), u(\sigma_k)]$$

where F is a function to be found out.

We assume that the quantitative – qualitative information $i_{pu}(\sigma_k)$ provided by two events independent both statistical and causal is equal to the sum of two pieces of information attached to each of them. Therefore,

$$(8)\ i_{pu}(\sigma_k) = i_{pu}(m_i) + i_{pu}(m_j)$$

Considering (7), (8) becomes

$$(9)\ F[p(\sigma_k), u(\sigma_k)] = F(p_i, u_i) + F(p_j, u_j)$$

In order to determine the solution of the functional equation (9), the following functions are defined

$$(10)\ z_i = \log_2 p_i\ ,$$

$$(11)\ z_j = \log_2 p_j$$

Using (10) and (11), (9) becomes

$$(12)\ F(2^{z_i+z_j}, u_i + u_j) = F(2^{z_i}, u_i) + F(2^{z_j}, u_j)$$

Denoting

$$(13)\ G(z,u) = F(2^z, u)\ ,$$

equation (12) can be written as

$$(14)\ G(z_i + z_j, u_i + u_j) = G(z_i, u_i) + G(z_j, u_j)$$

For $z_i = z_j = 0$, we have

$$(15)\ G(0, u_i + u_j) = G(0, u_i) + G(0, u_j)$$

Denoting

$$(16)\quad f(u)=G(0,u)\,,$$

equation (15) becomes

$$(17)\quad f(u_i+u_j)=f(u_i)+f(u_j)\,.$$

If the function $f(u)$ is continuous at least in one point, the solution of the functional equation (17) is (Fihtengol’c, 1964),

$$(18)\quad f(u)=a\cdot u,\ (\forall)\ u,\ a\in R\,.$$

For $u_i=u_j=0$, from (14) we obtain

$$(19)\quad G(z_i+z_j,0)=G(z_i,0)+G(z_j,0)$$

Denoting

$$(20)\quad g(z)=G(z,0)$$

equation (19) becomes

$$(21)\quad g(z_i+z_j)=g(z_i)+g(z_j)\,.$$

If the function $g(z)$ is continuous at least in one point, the solution of the functional equation (21) is

$$(22)\quad g(z)=b\cdot z,\ (\forall)\ z,\ b\in R\,.$$

For $u_i=z_j=0$, from (14) we have

$$(23)\quad G(z_i,u_j)=G(z_i,0)+G(0,u_j)$$

Making use of (16) and (20), we can write

$$(24)\quad G(z,u)=g(z)+f(u)$$

With (18) and (22), (24) becomes

$$(25)\quad G(z,u)=b\cdot z+a\cdot u$$

or, considering (11) and (13), we obtain

$$(26)\quad F(p_j,u_j)=b\cdot\log_2 p_j+a\cdot u_j$$

In the following we set $b=-1$ and $a=1$ and hence the quantitative - qualitative information attached to the message m_j is

$$(27)\quad i_{pu}(m_j)=-\log_2 p_j+u_j\,.$$

The quantitative – qualitative information $i_{pu}(m_j)$ defined in (27) determines a discrete random variable which takes on values with probabilities $p_j,\ j=1,2,...,n$. The average value of this information, denoted by $H_{pu}(S)$, can be computed by (3).

Theorem 2 - The maximum value of the entropy given in (3) for given utilities, is obtained when

$$(28)\quad p_j=\frac{2^{u_j}}{\sum_{j=1}^{n}2^{u_j}},\ j=1,2,...,n\,.$$

Proof: We consider the function

$$(29)\quad \begin{aligned}&\phi(p_1,p_2,...,p_n;u_1,u_2,...,u_n)=\\&-\sum_{j=1}^{n}p_j\log_2 p_j+\sum_{j=1}^{n}p_ju_j+\lambda(\sum_{j=1}^{n}p_j-1)\end{aligned},$$

where λ is a real positive number (the Lagrange multiplier).

The extreme of the function ϕ with respect to p_j for u_j fixed, coincides with the extreme of the function $H_{pu}(S)$. The necessary condition of extreme is given by the system

$$(30)\quad \begin{cases}\dfrac{\partial\phi(p_1,p_2,...,p_n;u_1,u_2,...,u_n)}{\partial p_j}=0,1\le j\le n\\ \sum_{j=0}^{n}p_j=1\end{cases}$$

From (30), (28) results. This extreme is a maximum one, because

$$(31)\quad \frac{\partial^2\Phi}{\partial p_j^2}=-\frac{\sum_{j=1}^{n}2^{u_j}}{2^{u_j}\ln 2}<0$$

and

$$(32)\quad \frac{\partial^2\Phi}{\partial p_j\partial p_i}=0,\ i\neq j$$

Theorem 3 - The maximum value of $H_{pu}(S)$ is:

$$(33)\quad \max_{p_j}H_{pu}(S)=\log_2\left(\sum_{j=1}^{n}2^{u_j}\right)$$

Proof: Relation (33) follows straightforward, by replacing (28) into (3). Obviously, for the probabilities p_j to sum up to 1, $p_n=p_{n-1}$.

3. BINARY HUFFMAN ENCODING FOR QUALITATIVE SOURCES

Once the probabilities p_j have been determined by (28), the average length of the code words will become minimum, if we use classical binary encoding procedures, as the Huffman one.

For convenience, we consider that the following relations are fulfilled:

$$(34)\quad \sum_{j=i+1}^{n} 2^{u_j} \le 2^{u_i};\; i=1,2,...,n-2 .$$

$$(35)\quad u_1 \ge u_2 \ge u_3 \ge \ldots \ge u_{n-1} = u_n .$$

For a binary code alphabet, $X=\{0,1\}$, according to the Huffman encoding procedure, Fig. 1 results. In this figure, the probabilities p_j are those given in (28), with $p_{n-1}=p_n$.

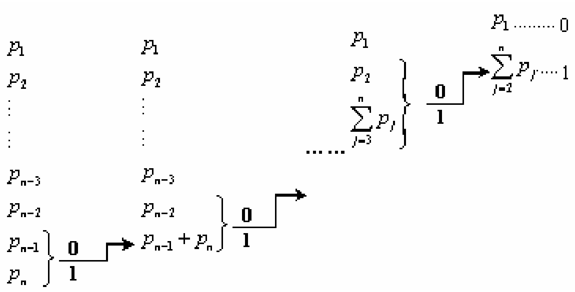

Fig. 1. The binary Huffman encoding procedure

The tree graph attached to the obtained code is given in Fig. 2.

Theorem 4 **-** The average length of the code words is

$$(36)\quad \bar{l} = 1 + \frac{2^{u_2}+2\cdot 2^{u_3}+...+(n-3)\cdot 2^{u_{n-2}}+(n-2)\cdot 2^{u_{n-1}+1}}{\sum_{j=1}^{n} 2^{u_j}}$$

Proof: The average length of the code words is equal to the sum of the intermediate node probabilities, inclusive the root, in the tree graph, that is

$$(37)\quad \bar{l} = 1+\sum_{i=2}^{n}\frac{2^{u_i}}{\sum_{j=1}^{n}2^{u_j}}+\sum_{i=3}^{n}\frac{2^{u_i}}{\sum_{j=1}^{n}2^{u_j}}+...+\sum_{i=n-1}^{n}\frac{2^{u_i}}{\sum_{j=1}^{n}2^{u_j}}$$

But

$$(38)\quad \sum_{j=1}^{n} 2^{u_j} = K = const.$$

Considering (37) and (38), (36) follows.

As the average length of the code words is obtained by summing the intermediate node probabilities, inclusive the root in the tree graph, it can be calculated, in general, without the constraints (34) and (35).

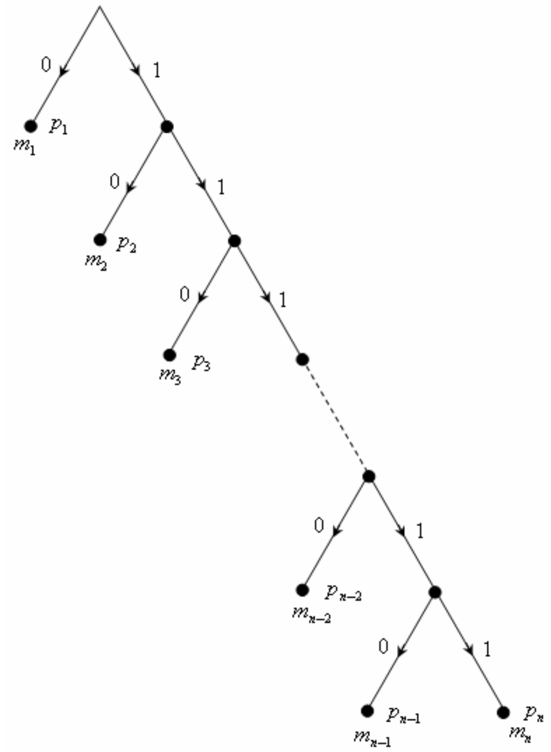

Fig. 2. The tree graph attached to the binary Huffman code

4. CONCLUSIONS

In this paper a procedure to encode qualitatively characterized sources on noiseless channels is proposed. In this case the messages the source delivers are accompanied by certain real numbers which characterize the quality (cost, importance or utility) from the user’s point of view.

As the message delivering probabilities are unknown, the optimal encoding procedures as Shannon Fano, Huffman etc. cannot be applied. To apply these optimal procedures, we derive the probabilities with which the messages should be delivered, so that the mean information per code word is maximum one. To this aim the entropy is derived for a source that delivers the messages $m_j,\ j=1,2,...,n$, with probabilities p_j and utilities u_j. Then, the maximum value of this entropy is derived with respect to the probabilities p_j for imposed utilities. Thus, the probabilities $p_j\ \ j=1,2,...,n$, are expressed in terms of the utilities. (The entropy computed in this paper has the same main properties as the Shannon’s entropy).

Once determined the probabilities p_j with which the source messages have to be transmitted, so that the entropy is maximum for imposed message utilities, we can use any classical optimal encoding procedures on noiseless channels. These methods can be applied for both binary and nonbinary code alphabet.

5. REFERENCES

Huffman, D. A. (1953). *A method for the construction of minimum redundancy codes.* Butterwoths Scientific Publications London Ed. Jackson, Communication Theory, pp. 102-110.

Gallager, G. R. (1968). *Information theory and reliable communications.* New York: John Wiley and Sons Inc..

Cover T. and J. A. Thomas, (1991). *Elements of information theory.* New York: John Wiley and Sons Inc.

Belis M. and S. Guiasu. (1968). A quantitative-qualitative measure of information in cybernetics., *IEEE Trans. Inform. Theory*, IT – **14**, pp. 593-594.

Khan A. B. and R. Autar. (1979). On useful information of order α and β. *Soochow J. Math.*, **5**, pp. 93-99.

R. Autar and A. B. Khan, "On generalized useful information for incomplete distribution", *J. of Comb. Information and Syst. Sci.,* **14**, no.4, pp. 187-191, 1989.

Gurdial A. and F. Pessoa. (1977). On useful information of order α. *J Comb. Information and Syst. Sci.*, **2** , pp. 158-162.

Munteanu V. and P Cotae. (1992). The entropy with preference. *Int. J. Electronics and Comm. AEU*, **46**, pp. 429 – 431.

Longo, G. (1976). A noisless coding theorem for sources having utilities. *SIAM J. Appl. Math.*, **30**, No.4, pp. 739-748.

Khan, A. B., B. A. Bhat and S. Pirzada. (2005). Some results on a generalized useful information measure. *J. Inequal. Pure and Appl. Math*, **6**, No. 4, pp.1-5.

Shannon, C. E. (1948). A mathematical theory of communication. *BSTJ*, **27**, pp. 379-423, 623-656.

Fihtengol'c, G. M.(1964). *Differential and integral calculus*. Ed. Tehnica, Bucharest.