Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 51(65), Fascicola 2, 2006

Some Properties of Semantic Sources

Valeriu Munteanu[1], Daniela Tarniceriu[1]

Abstract – In this paper we derive the quantitative – qualitative entropy an extension of order *m* of a semantic source, as well as the semantic entropy of discrete ergodic sources with memory. The Kraft inequality and Shannon's first theorem are generalized for these sources. Some applications of semantic sources are also presented.
Keywords: **semantic sources, entropies, Kraft's inequality, Shannon's first theorem.**

I. INTRODUCTION

Unlike the quantitative characterization of information [1,2,3], in [4] we have introduced the concept of quantitative – qualitative information, and derived the entropy for semantic sources. The main properties of entropy for these sources are also established. Longo [5], Guardial and Pessoa [6], Khan and Autar [7], Autar and Khan [8], Khan and Bhat [9] have studied generalized coding theorems by considering different generalized measures of information. In this paper we determine the entropy for an extension of a semantic source, and then we consider semantic sources with memory to determine their entropy. The Kraft inequality [3] and the first Shannon theorem [2] are extended for semantic sources. In the last part of our work we present some applications of semantic entropy.

II. DETERMINING THE ENTROPY FOR AN EXTENSION OF A SEMANTIC SOURCE

For many source models it is useful to consider that the source delivers groups of messages, instead individual ones. Generally, from a discrete, complete and memoryless source S which delivers *n* messages, $s_1, s_2, ..., s_n$, with probabilities $p_1, p_2, ..., p_n$ and utilities $u_1, u_2, ..., u_n$, characterized by the distribution

$$S: \begin{pmatrix} s_1 & s_2 & \cdots & s_n \\ p_1 & p_2 & \cdots & p_n \\ u_1 & u_2 & \cdots & u_n \end{pmatrix} \quad (1)$$

We can derive another source, called the extension of the order *m* of the former, denoted by S^m, consisting in groups of *m* messages of the source *S,* in all possible combinations .

The extension of order *m* contains n^m composite symbols:

$$\sigma_k = s_{k1}s_{k2}s_{k3}...s_{km},\ k = 1,2,...,n^m \quad (2)$$

where $s_{k1}, s_{k2}, s_{k3}, ..., s_{km}$ are messages of the source *S*.

Assuming *S* memoryless, and its messages independent both probabilistic and logic – causal, we have

$$p(\sigma_k) = p(s_{k1}) \cdot p(s_{k2}) \cdot ... \cdot p(s_{km}) \quad (3)$$

and

$$u(\sigma_k) = u(s_{k1}) + u(s_{k2}) + ... + p(s_{km}) \quad (4)$$

Theorem 1

The quantitative – qualitative entropy of the extension of order *m* is *m* times the entropy of the semantic source, that is

$$H_{pu}(S^m) = mH_{pu}(S) \quad (5)$$

Proof

We prove this theorem by induction. First, we verify easily that (5) is true for $m = 1,2$. Next, we assume that (5) holds true for *m* and prove that

$$H_{pu}(S^{m+1}) = (m+1)H_{pu}(S) \quad (6)$$

Let ξ_i be a composite symbol made of $m+1$ messages of the source *S*, as

$$\xi_i = \sigma_k s_j,\ k = 1,2,...,n^m;\ j = 1,2,...,n \quad (7)$$

where s_j is a messages the source S delivers.

Since

$$p(\xi_i) = p(\sigma_k)p(s_j),\ u(\xi_i) = u(\sigma_k) + u(s_j) \quad (8)$$

$$\sum_{k=1}^{n^m} p(\sigma_k) = 1,\ \sum_{j=1}^{n} p(s_j) = 1, \quad (9)$$

the entropy of the extension of order (*m*+1) becomes

$$\begin{aligned} &H_{pu}(S^{m+1}) = \\ &-\sum_{i=1}^{n^{m+1}} p(\xi_i)\log_2 p(\xi_i) + \sum_{i=1}^{n^{m+1}} p(\xi_i)u(\xi_i) = \\ &-\sum_{k=1}^{n^m} p(\sigma_k)\log_2 p(\sigma_k) + \sum_{k=1}^{n^m} p(\sigma_k)u(\sigma_k) - \\ &-\sum_{j=1}^{n} p(s_j)\log_2 p(s_j) + \sum_{j=1}^{n} p(s_j)u(s_j) \end{aligned} \quad (10)$$

which is equivalent to (6).

[1] Facultatea de Electronică şi Telecomunicaţii, Departamentul Comunicaţii Bd. Carol I, Nr. 11, 700506 Iasi, e-mail vmuntean@etc.tuiasi.ro

III. DETERMINING THE ENTROPY FOR A SEMANTIC SOURCE WITH MEMORY

Let us consider a discrete, complete and ergodic source with memory of order *m*, with the distribution given in (1).

Theorem 2

The entropy of sources with memory of order *m* is

$$H_{pu}^{m}(S) = -\sum_{i=1}^{n^m} \sum_{j=1}^{n} p_i \cdot p(s_j \mid S_i) \cdot \log_2 p(s_j \mid S_i) + \sum_{i=1}^{n^m} \sum_{j=1}^{n} p_i \cdot p(s_j \mid S_i) \cdot u(s_j \mid S_i) \quad (11)$$

Proof

Let S_i be the state characterized by the sequence

$$S_i \rightarrow s_{i1}, s_{i2}, \ldots, s_{im} . \quad (12)$$

We denote

$$p(s_j \mid s_{i1}, s_{i2}, \ldots, s_{im}) \overset{not.}{=} p(s_j \mid S_i) \quad (13)$$

the probability that the source delivers the message s_j, given the state S_i, and

$$u(s_j \mid s_{i1}, s_{i2}, \ldots, s_{im}) \overset{not.}{=} u(s_j \mid S_i) \quad (14)$$

the utility the message s_j possesses, given the state S_i.

In [4], we proved that for a memoryless semantic source, the information attached to the message s_k having the probability p_k and the utility u_k, is given by

$$i_{pu}(s_k) = -\log_2 p_k + u_k . \quad (15)$$

Then, the quantitative – qualitative information obtained when the source is in the state S_i and it delivers the message s_j is then given by

$$i_{pu}(s_j \mid s_{i1}, s_{i2}, \ldots, s_{im}) = -\log_2 p(s_j \mid S_i) + u(s_j \mid S_i) \quad (16)$$

From the state S_i any message s_j can be delivered with a certain conditional probability (even equal to zero, if from that state a certain message cannot be delivered). The average quantitative – qualitative information the state S_i can deliver is

$$i_{pu}(S_i) = \sum_{j=1}^{n} p(s_j \mid S_i) i_{pu}(s_j \mid S_i) \quad (17)$$

or, considering (16)

$$i_{pu}(S_i) = -\sum_{j=1}^{n} p(s_j \mid S_i) \log_2 p(s_j \mid S_i) + \sum_{j=1}^{n} p(s_j \mid S_i) u(s_j \mid S_i) \quad (18)$$

Denoting by p_i, $i = 1, 2, \ldots, n^m$, the state probabilities of the ergodic, discrete, complete source with memory, the average quantitative – qualitative information, or the quantitative – qualitative entropy, denoted by $H_{pu}^{m}(S)$ can be computed by

$$H_{pu}^{m}(S) = \sum_{i=1}^{n^m} p_i i_{pu}(S_i) \quad (19)$$

Considering (18) and (19), we get (11).

IV. KRAFT'S THEOREM FOR SEMANTIC SOURCES

Theorem 3

The Kraft's theorem for semantic sources is

$$\sum_{k=1}^{n} M^{-l_k} \cdot 2^{u_k} \leq 1 , \quad (20)$$

where n is the number of messages the information source supplies, M – the number of symbols in the code alphabet and l_k - the length of the codeword c_k, $(\forall) k = 1, 2, \ldots, n$.

Proof

Let *S* be the semantic source characterized by the distribution given in (1).

Let

$$X = \{x_1, x_2, \ldots, x_M\} \quad (21)$$

be the alphabet of the code,

$$C = \{c_1, c_2, \ldots, c_n\} \quad (22)$$

the code words attached to the messages, and

$$L = \{l_1, l_2, \ldots, l_n\} \quad (23)$$

the length of the code words.

Due to the one – to one correspondence between the messages $s_k \in S$ and the code words $c_k \in C$, the information attached to the message s_k is equal to that attached to the code word c_k, i. e.

$$i_{pu}(s_k) = i_{pu}(c_k) \quad (24)$$

On the other hand, the maximum information per symbol of the code alphabet is $H_{\max}(X) = \log_2 M$, which can be reached when the symbols of the code alphabet are used independently and equally likely.

The length l_k corresponding to the code word c_k has to satisfy

$$l_k \geq \frac{i_{pu}(c_k)}{\log_2 M} = \frac{-\log_2 p_k + u_k}{\log_2 M} \quad (25)$$

From (24) and (25), we have

$$p_k \geq 2^{u_k} \cdot M^{-l_k} . \quad (26)$$

As

$$\sum_{k=1}^{n} p_k = 1 \quad (27)$$

eq. (20) results.

V. SHANNON'S FIRST THEOREM FOR SEMANTIC SOURCES

Let there be the source characterized by (1) and the code characterized by (21), (22) and (23).

Obviously, the probabilities and utilities of the source messages are equal to the probabilities $p(c_k)$ and utilities $u(c_k)$ of the code words, respectively, i. e.

$$p_k = p(c_k) \quad (28)$$

$$u_k = u(c_k) \quad (29)$$

The length of each code word must belong to the set of positive integers, therefore l_k must be chosen as an integer satisfying the condition

$$\frac{-\log_2 p_k + u_k}{\log_2 M} \le l_k < \frac{-\log_2 p_k + u_k}{\log_2 M} + 1 \quad (30)$$

where $\log_2 M$ is the maximum value $H(X)$ can take on.

By multiplying (30) by p_k and summing up from 1 to n, we have

$$\frac{-\sum_{k=1}^{n} p_k \log_2 p_k + \sum_{k=1}^{n} p_k u_k}{\log_2 M} \le \sum_{k=1}^{n} p_k l_k < \frac{-\sum_{k=1}^{n} p_k \log_2 p_k + \sum_{k=1}^{n} p_k u_k}{\log_2 M} + 1 \quad (31)$$

or,

$$\frac{H_{pu}(S)}{\log_2 M} \le \bar{l} < \frac{H_{pu}(S)}{\log_2 M} + 1 \quad (32)$$

where

$$H_{pu}(S) = -\sum_{k=1}^{n} p_k \log_2 p_k + \sum_{k=1}^{n} p_k u_k \quad (33)$$

is the entropy of the semantic source [4] and

$$\bar{l} = \sum_{k=1}^{n} p_k l_k \quad (34)$$

is the average length of the code words. Relation (32) holds true also for the extension of order m, for which we can write

$$\frac{H_{pu}(S^m)}{\log_2 M} \le \bar{l}_m < \frac{H_{pu}(S^m)}{\log_2 M} + 1 \quad (35)$$

where $\bar{l}_m$ is the average length of the code words corresponding to a sequence of m messages of the source S and $H_{pu}(S^m)$ is its entopy.

Since

$$\bar{l} = \frac{\bar{l}_m}{m}, \quad (36)$$

we have

$$\frac{H_{pu}(S)}{\log_2 M} \le \bar{l} < \frac{H_{pu}(S)}{\log_2 M} + \frac{1}{m} \quad (37)$$

From (37) it follows that, when $m \to \infty$, the average length of the code words becomes equal to the minimum average length. Thus, (37) becomes a generalization of C. E. Shannon's Theorem [2], for encoding discrete sources for noiseless channels, in the case of cybernetics systems. If the utilities of the messages of source S are zero, the classical results are obtained [1].

VI. APPLICATIONS OF QUALITATIVE – QUANTITATIVE ENTROPY

A first application of quantitative – qualitative entropy regards the calculus of the entropy for a binary block code.

Let us consider a binary block code, for which each codeword contains N binary symbols. If k denotes the number of information symbols in each codeword, the number of code words, n, is determined by

$$n = 2^k \quad (38)$$

Due to the one - to – one correspondence between the code words and the messages of the information source, each codeword will have the same probability and utility as the corresponding message.

Let us also suppose that in order to correct the errors in each codeword, m parity – check symbols are used.

Considering the utility of each transmitted codeword equal to the number of parity check symbols m, the following distribution results:

$$C : \begin{pmatrix} c_1 & c_2 & \cdots & c_n \\ m & m & \cdots & m \end{pmatrix} \quad (39)$$

The absolute maximum entropy $H_{ma}(C)$ of this source is attained when the messages and the code words are equally likely delivered.

From [4] we have

$$H_{ma}(C) = \log_2 n + \frac{U}{n} \quad (40)$$

Considering (38) and the fact that the whole utility of the code words is

$$U = n \cdot m, \quad (41)$$

we get

$$H_{ma}(C) = k + m = N \quad (42)$$

According to (42) the absolute maximum entropy of a binary block error correcting code is equal to the codeword length.

A second application consists in the establishing of the delivering probabilities of unequal protected code words, so that the average information per codeword is maximum one. With this purpose in view, let us consider a binary block code of length N, in which the first codeword has m_1 parity check symbols, the second one, m_2, and so on, the last one having m_n parity check symbols.

Further on, we consider the general case, in which code words with the same number of parity check symbols could exist. Obviously, the more parity check symbols the code words contain, the more errors can be corrected.

We also consider that the parity check symbols in each codeword represent the utilities.

In order to obtain the maximum average information per codeword, the source characterized by the distribution

$$S : \begin{pmatrix} s_1 & s_2 & \cdots & s_n \\ m_1 & m_2 & \cdots & m_n \end{pmatrix} \quad (43)$$

provides its messages with the probabilities computed by [4]

$$p_k = \frac{2^{m_k}}{\sum_{j=1}^{n} 2^{m_j}}, k = 1, 2, ..., N \quad (44)$$

The average information per codeword may be computed by means of [4]

$$H_m(S) = \log_2 \left(\sum_{k=1}^{n} 2^{m_k} \right) \quad (45)$$

The average information per symbol in a codeword results as follows:

$$i = \frac{H_m(S)}{N} = \frac{\log_2 \left(\sum_{j=1}^{n} 2^{m_j} \right)}{N} \quad (46)$$

where n is given by (38).

In the particular case, when all code words are identically protected

$$m_1 = m_2 = ... = m_N = m \,, \quad (47)$$

an information $i = 1$ bit/symbol is obtained.

The third application consists in the implementation of an encoding method for noiseless channels of sources characterized by the distribution

$$S : \begin{pmatrix} s_1 & s_2 & \cdots & s_n \\ u_1 & u_2 & \cdots & u_n \end{pmatrix}, u_k \in R, k = 1, 2, ..., n \,, \quad (48)$$

so that the average information per codeword is maximum and the average codeword length is minimum.

From (48) we observe that each message utility is known and we want to find the message delivering probabilities, so that the average information per message is maximum one. To this aim the probabilities p_k with which the messages s_k have to be delivered are computed by means of (44). Therefore, the source entropy becomes maximum one. In order to obtain the code words of the smallest length, for a noiseless channel, one can use the Huffman encoding procedure [12], using the above obtained probabilities.

VII. CONCLUSIONS

In this paper, we derive the quantitative – qualitative entropies of an extension of order m of the source (eq. 5) and of a discrete ergodic source with memory (eq. 11). These relations represent generalizations of the classical concepts on information. The quantitative – qualitative information results as the sum between a quantitative information and a qualitative one. If the qualitative characteristic is neglected, the classical known relations [1] are obtained. When only the qualitative characteristic is required, the first term in the relations above is dropped out.

By extending the notion of entropy to cybernetic systems, the Kraft inequality and Shannon's first theorem have been generalized. Three applications of sources with preferences are also presented. The first one regards the error correcting block codes, the second one, the unequally error protection block codes and the third one, the possibility to encode sources characterized only qualitatively, so that the average information per codeword is maximum and the average length of code words is minimum.

REFERENCES

[1] G. R.Gallager, *Information theory and reliable communications*, New York, John Wiley and Sons Inc., 1968.

[2] C. E.Shannon, "A mathematical theory of communication", BSTJ, 27 1948, pp. 379-423, 623-656.

[3] T.Cover, J. A.Thomas, *Elements of Information Theory*, Wiley, 1991.

[4] V.Munteanu, D. Tarniceriu, "On Semantic Feature of Information", Symposium Etc. '06", *Buletinul Universității "Politehnica", Seria Electrotehnica, Electronica si Telecomunicatii*, Tom 51 (65), 2006, Fascicola 1-2, 2006.

[5] G. Longo, "A noisless coding theorem for sources having utilities", *SIAM J. Appl. Math*., 30 (4), 1976, pp. 739-748.

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

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

[8] R. Autar, A. B. Khan, "On generalized useful information for incomplete distribution", *J. of Comb. Information and Syst. Sci*., 14 (4) (1989), 187-191.

[9] A. B. Khan, B. A. Bhat, S. Pirzada, "Some results on a generalized useful information measure". *J. Inequal. Pure and Appl. Math*, 6 (4), 2005, pp. 1-5.

[10] M. Belis, S. Guiasu, "A quantitative-qualitative measure of information in cybernetics", *IEEE Trans. Inf. Theory* IT – 14, 1968, pp. 593-594.

[11] V. Munteanu, P. Cotae, "The entropy with preference", *Int. J. Electronics and Comm. AEU*, 46, 1992, pp. 429 – 431.

[12] D. A. Huffman, "A Method for the construction of minimum redundancy codes", *Ed. Jackson, Communication Theory. Butterwoths Scientific Publications London*. 1953 pp. 102-110.