Encoding of Information generated by an unknown Source without Memory with infinite Alphabet

Viktor K. Trofimov[1], Tatiana V. Khramova[1]
[1]Siberian state university of telecommunications and information sciences, Novosibirsk, Russia

***Abstract* – The problem of encoding a source with an infinite, countable alphabet arises, for example, when compressing a sequence obtained by quantizing the results of orthogonal transformations of a random process [2]. Encoding positive integers it is also used in codes, Ziv-Lempel [3].**

***Index Terms* – entropy, redundancy, weakly universal coding.**

I. Introduction

LET letters of the infinite countable alphabet

$$A = \{a_1, a_2, \ldots, a_n, \ldots\},$$

hereinafter referred to as the input, be generated independently by the source θ, and the probability of generating the letter a_i by the source θ is θ_i $i = 1, 2, \ldots$. The sequence of letters generated by the source is divided into blocks (words) of length n and encoded by the words of the output alphabet B, which, without reducing the generality, can be considered binary.

If u – is an arbitrary word of length n, we denote $P_\theta(u)$ the probability of this word, provided that it is generated by the source θ, in particular

$$P_\theta(a_i) = \theta_i, i = 1, 2, \ldots.$$

Let $\phi(u)$ a code word for u by encoding ϕ, and $|\phi(u)|$ - the length of the code word $\phi(u)$. By *word's length* we mean the number of letters of the corresponding alphabet in the word. *The entropy of the source* θ is the value $H(\theta)$ determined by the equality

$$H(\theta) = -\sum_{i=1}^{\infty} \theta_i \log \theta_i .$$

Here and on $\log x = \log_2 x$, $0 \log 0 = 0$.

Denote by Ω set of sources with finite entropy, $\theta \in \Omega$ then and only then, when $H(\theta) < \infty$. Following [3,5,9], the coding ϕ efficiency will be estimated by its *redundancy* $R(n, \phi, \theta)$ determined by the equality

$$R(n, \phi, \theta) = \frac{1}{n} \sum_{|u|=n} P_\theta(u) |\phi(u)| - H(\theta).$$

According to the basic theorem of K. Shannon, any encoding ϕ always performs an inequality

$$R(n, \phi, \theta) \geq 0 .$$

At a fixed source θ, as follows from [5], there is an encoding ϕ_0 such that

$$0 < R(n, \phi, \theta) < \frac{1}{n},$$

i. e. redundancy of coding of the known source with the counting alphabet, as well as for the source generating finite number of letters, tends to zero with speed $o(n^{-1})$.

Consider a situation where a source θ is unknown but is known to belong to a source class Ω.

Value $R(n, \Omega)$, defined by the equation

$$R(n, \Omega) = \inf_{\phi} \sup_{\theta \in \Omega} R(n, \phi, \theta),$$

following [6,9], we call *redundancy of universal coding* for a set of sources Ω.

In the case of a finite input alphabet, universal coding was studied by many authors, both here and abroad, a detailed bibliography can be found in [9].

II. Problem Definition

Let's determinate coding ϕ *weakly universal* for a set of sources Ω, if for any $\theta \in \Omega$ equality

$$\lim_{n \to \infty} R(n, \varphi, \theta) = 0$$

is always executed, while

$$\lim_{n \to \infty} R(n, \Omega) \neq 0 .$$

For the first time, weakly universal coding was considered in [8] for a set of all stationary sources. It should be noted that coding with counting alphabet for the first time has been studied in sufficient detail in [10].

In this paper, we studied the coding of Poisson sources, in the presence of a sample generated by the source. In [10], it was proposed to encode, which at low sample volumes is far from optimal, and at samples comparable to the length of the encoded block, gives the order of decreasing redundancy, the same as for the known statistics.

In work [11] the optimal coding of a set of all Poisson sources with unknown parameter λ is offered. It is proved that if value λ is limited, there is a universal coding, and the offered coding is asymptotically optimum. If parameter λ is not limited, only weakly universal encoding exists. In addition, if you have a selection of any size other than 0, you are offered an optimal encoding that does not depend on the parameter λ.

Obviously, there is no universal coding for set of sources Ω with countable alphabet and finite entropy. It is also not possible to obtain estimates of weakly universal coding for multiple sources. Therefore, in [1] two classes of sources were studied, for which weakly universal coding methods were proposed and their efficiency, characterized by the rate of reduction of redundancy, was estimated. Two sets of sources Ω_q and $\Omega_{\exp}$ with limited central moments are considered in [1]. Source $\theta \in \Omega_q$ if inequality

$$\sum_{j=1}^{\infty} j^{\theta} \theta_q = m_q < \infty$$

is satisfied.

In [1], a coding method ϕ_q is proposed such that an asymptotic inequality

$$R\left(n, \phi_q, \theta\right) \lesssim \frac{A\left(m_q\right)^{\frac{1}{q+1}}}{n^{\frac{q}{q+1}}}$$

is satisfied for any θ, $\theta \in \Omega_q$. Here and on $f(n) \gtrsim g(n)$, if

$$\lim_{n \to \infty} f(n) / g(n) = \infty.$$

Source θ in $\Omega_{\exp}$, if inequality

$$\sum_{j=1}^{\infty} 2^j \theta^j < \infty$$

is satisfied.

In [1] a method of encoding $\phi_{\exp}$ is proposed such that for any θ, $\theta \in \Omega_{\exp}$ inequality

$$R\left(n, \phi_{\exp}, \theta\right) \lesssim \frac{1}{2} \frac{\log^2 n}{n}$$

is satisfied.

III. THEORY

In this paper, for these source classes, we obtain lower bounds on the redundancy of universal coding, in descending order coinciding with the upper ones, and thereby prove the asymptotic optimality of the species proposed in [1].

Theorem 1. If Ω_q - a set of discrete independent sources with a countable alphabet with bounded moment q that there is a subset of sources $\tilde{\Omega}_q$, and the constant c_1 such that when $n \to \infty$ an asymptotic inequality

$$R\left(n, \tilde{\Omega}_q\right) \gtrsim \frac{c_1}{n^{\frac{q}{q+1}}}$$

is satisfied.

Theorem 2. If $\Omega_{\exp}$ - a set of discrete independent sources with a countable alphabet and uniformly bounded moments, then there is a subset of sources $\tilde{\Omega}_{\exp}$, $\tilde{\Omega}_{\exp} \in \Omega_{\exp}$ and the constant c_2 such that when $n \to \infty$ the ratio

$$R\left(n, \tilde{\Omega}_{\exp}\right) \gtrsim c_2 \cdot \frac{\log^2 n}{n} .$$

For the proof of theorems 1 and 2, we construct sets $\tilde{\Omega}_q$, $\tilde{\Omega}_{\exp}$ which depend on the length of the encoded block , but at any fixed *n* are finite.

IV. DISCUSSION OF RESULTS

From theorems 1, 2 it follows that the decreasing rate of redundancy of universal coding for a set of sources Ω_q, $\Omega_{\exp}$ obtained in [1], asymptotically coincide.

In the proof of theorems 1, 2 the convergence rate

$$\left[H\left(v_t(\theta)\right) - H(\theta)\right]$$

is investigated, which is compared with the convergence rate of the value

$$R\left(n, \phi_0(t), v_t, (\theta)\right)$$

and the optimal ratio between these values is found.

Source $v_t(\theta)$ has an input alphabet containing t letters; the first $t-1$ letters coincide with the letters $a_1, a_2, \ldots, a_{t-1}$ of the input alphabet A, all other letters of the source θ are perceived as one letter a_t of the source $v_t(\theta)$. So, the source $v_t(\theta)$ is determined by the letters

$$A_t = \{a_1, a_2, \ldots, a_{t-1}, a_t\}$$

of the input alphabet, which are generated independently with probabilities

$$v_t(\theta) = \theta_i, \; i = 1, 2, \ldots, t-1,$$

$$v_t(\theta) = \sum_{i=1}^{\infty} \theta_i .$$

V. CONCLUSION

As follows from [1,2,10,11] and the results of this paper, for special classes of sources without memory, it is possible to construct codes that, at any source from these classes, make it possible to efficiently compress the information generated by these sources.

At the same time, the assessment of redundancy significantly depends on the source. For individual source classes, such as Poisson classes with a bounded parameter, it is possible to build an encoding that is efficient for all sources, and the encoding of redundancy is independent of the source being encoded.

REFERENCES

[1] *Babkin V.F.*. Universal coding of discrete independent sources with counting alphabet // Voprosy kibernetiki – Moskva, 1977. – No.34. Redundancy problems in information systems. – pp.3-10 (In Russian)

[2] *Bocharova I.* Compression for Multimedia. Cambridge: Cambridge Univ. Press, 2010.

[3] *Ziv J., Lempel A.* Compression of Individual Sequences via Variable-Rate Coding / / IEEE Trans. Inform theory. 1978. V. 24. № 5. P. 530-536

[4] *Shennon K.* Mathematical theory of communication. Works on information theory and Cybernetics. - 1969. – pp.243–332. (In Russian)

[5] *Levenshteyn V.I.*. Redundancy and deceleration of separable encoding of natural numbers // Problemy kibernetiki. – M., Nauka, 1968. – No.20 – pp.173-179 (In Russian)

[6] *Krichevskii R.E.* The relation between redundancy coding and the reliability of source information // Problemy Peredachi Informacii. 1968. – V.4.No 3.-pp.48-57 (In Russian)

[7] *Krichevskii R.E., Trofimov V.K.* The Performace of Universal Encoding // IEEE Trans. on Inform. Theory. 1981.- V.IT-27. №2. pp.199–207.

[8] *Shtarkov Yu.M., Babkin V.F.* Combinatorial encoding for discrete stationary sources // In 2-nd Inter. Symp. on Inform. Theory. Tsahkad-zor, Armenia, USSR. 1971. Academia – Kiado. Budapesht. – 1973. pp.249-256

[9] *Krichevskii R.E.* Universal Compression and Retrieval, - 1994. Kluwer Academic Publishers. Dordecht/Boston/London. P.219.

[10] *Shtarkov Yu.M., Rybeva N.A.* Adaptive encoding for Poisson sources. Coding in complex systems. - M.: Nauka 1974. – pp.169-175.

[11] *Trofimov V.K.*. Encoding Poisson streams with unknown and inaccurately known parameters. // Problemy Peredachi Informacii. – 1987. – V.23, № 4 – pp.49-57.

Trofimov Viktor Kupriyanovich

Siberian state university of telecommunications and information sciences. Doctor of technical sciences.

Birth date: 20.10.1949. Phone: +7-913-911-12-77, E-mail: trofimov@sibsutis.ru

Area of scientific research: information theory, data compressing, optimal coding.

Khramova Tatiana Viktorovna

Siberian state university of telecommunications and information sciences. Candidate of technical sciences.

Birth date: 21.06.1972. Phone: +7-913-929-41-00, E-mail: tvkhramova@gmail.com

Area of scientific research: information theory, data compressing, optimal coding.