

Кодирование геометрических источников при неизвестной статистике

В. К. Трофимов

Предложен метод универсального кодирования произвольного множества источников без памяти, порождающих буквы бесконечного алфавита. Вероятности появления букв входного алфавита являются геометрической прогрессией. Предложенный метод является слабоуниверсальным для множества всех геометрических источников. При знаменателе геометрической прогрессии превосходящей δ , $\delta > 0$, предложенное кодирование – универсальное. Получены оценки избыточности через ε -энтропию для произвольного подмножества геометрических источников.

Ключевые слова: кодирование, избыточность, энтропия, источник сообщений.

1. Введение

Предлагаемая статья посвящена универсальному кодированию источников с бесконечным алфавитом, имеющих геометрическое распределение вероятностей. В работе Левенштейна [1] доказано, что и при кодировании бесконечного числа сообщений выполнение неравенства Крафта является необходимым и достаточным условием существования дешифруемого кодирования. В дальнейшем мы будем рассматривать только дешифруемые в смысле [1] кодирования. В [2] установлено, что для кодирования блоков длины N , порожденных источником с бесконечным алфавитом, в общем случае не существует универсального кодирования.

Однако для определенных классов источников с бесконечным алфавитом были предложены так называемые методы слабого универсального кодирования [3]. Слабоуниверсальное кодирование – это кодирование, которое обеспечивает оптимальность для каждого источника из заданного множества источников, но не обеспечивает оптимальность для совокупности всех источников. Первым примером слабоуниверсального кодирования является предложенное Бабкиным и Штарьковым [4] кодирование множества всех стационарных источников с конечным алфавитом. Кодированию специальных классов источников с бесконечным входным алфавитом посвящены работы Бабкина [5] и Кудряшова, Порова [6]

Пуассоновские источники рассматривались Штарьковым и Рыбьевой [7], оптимальные коды для этих источников предложены автором в [8].

Настоящая работа посвящена универсальному кодированию геометрических источников, т.е. источников, в которых вероятности появления букв образуют геометрическую прогрессию. Используя идею построения универсальных кодов из [9, 10], предложен универсальный код для любого подмножества таких источников и получены оценки избыточности кодирования для рассматриваемых классов источников.

Доказано, что это кодирование является слабоуниверсальным для всех рассматриваемых источников и универсальным, если знаменатель прогрессии больше или равен δ , $0 < \delta < 1/2$

2. Основные определения

Предположим, что буквы бесконечного алфавита $A = \{a_0, a_1, a_2, \dots\}$, называемого в дальнейшем входным, порождаются источником без памяти θ . Вероятность порождения буквы a_j обозначим $P_\theta(a_j) = \theta_j$, $j = 0, 1, 2, \dots$. Очевидно, $\theta_0 + \theta_1 + \theta_2 + \dots = 1$.

Таким образом, каждый источник θ однозначно определяется последовательностью положительных чисел $\theta_0, \theta_1, \theta_2, \dots$, сумма которых равна 1, верно и обратное утверждение.

Как обычно, A^N – совокупность всех слов, содержащих N букв алфавита A . Рассмотрим произвольную полубесконечную последовательность букв алфавита A , порождаемую источником θ . Такая последовательность однозначно разбивается на слова (блоки) по N букв в каждом. Число букв в слове u , взятом в произвольном алфавите, называют длиной слова u и обозначают $|u|$.

Каждое из слов u , $u \in A^N$ с помощью отображения φ кодируется словом $\varphi(u)$ в выходном алфавите B , который, не уменьшая общности, можно считать двоичным, т.е. $B = \{0, 1\}$. Как отмечалось выше, мы рассматриваем только дешифруемые в смысле [1] кодирования.

В [1] установлено, что дешифруемое кодирование φ длинами кодовых слов $|\varphi(u)|$, $u \in A^N$ существует тогда и только тогда, когда выполняется неравенство:

$$\sum_{u \in A^N} 2^{-|\varphi(u)|} \leq 1. \quad (1)$$

Неравенство (1) является обобщением неравенства Крафта [11] на бесконечный алфавит. Среднее число букв выходного алфавита, приходящихся на одну букву входного при кодировании φ , назовём стоимостью кодирования и обозначим $c(N, \varphi, \theta)$:

$$c(N, \varphi, \theta) = \frac{1}{N} \sum_{u \in A^N} P_\theta(u) \cdot |\varphi(u)|. \quad (2)$$

Здесь и в дальнейшем $P_\theta(u)$ – вероятность порождения слова u источником θ . Если источник θ задается вероятностями θ_j , $j = 0, 1, 2, \dots$, то его энтропия $H(\theta)$ вычисляется по формуле [11]:

$$H(\theta) = - \sum_{j=0}^{\infty} \theta_j \log \theta_j. \quad (3)$$

По определению полагаем: $\log x = \log_2 x$, $0 \log 0 = 0$.

Разность между стоимостью кодирования $c(N, \varphi, \theta)$ и энтропией источника $H(\theta)$, которые определены равенствами (2), (3), обозначим $r(N, \theta, \varphi)$ и назовём избыточностью кодирования φ для источника θ при кодировании блоков длины N или просто избыточностью кодирования φ . Таким образом, по определению:

$$r(N, \theta, \varphi) = c(N, \varphi, \theta) - H(\theta). \quad (4)$$

Пусть Ω – произвольное множество источников. Наша задача состоит в том, чтобы предложить кодирование, которое будет хорошим для любого источника из Ω . Эффективность кодирования характеризует величина $R(N, \Omega)$, называемая избыточностью универсального кодирования для множества источников Ω , которая задается равенством

$$R(N, \Omega) = \inf_{\varphi} \sup_{\theta \in \Omega} r(N, \theta, \varphi), \quad (5)$$

здесь \inf берётся по всем дешифруемым кодированиям.

Кодирование, при котором $r(N, \theta, \varphi)$ с ростом N стремится равномерно к нулю, называют универсальным. Очевидно, что в этом случае выполняется неравенство:

$$R(N, \theta) \leq \sup_{\theta \in \Omega} r(N, \theta, \varphi). \quad (6)$$

Если же величина $r(N, \theta, \varphi)$ стремится к нулю неравномерно, то кодирование называют слабоуниверсальным.

В настоящей работе для множества всех геометрических источников предложено кодирование, которое является слабоуниверсальным и универсальным при $1 > \beta > \delta > 0$. Получена оценка избыточности предложенного кодирования для произвольного подмножества геометрических источников.

Доказано, что если $\tilde{\Omega}$ – произвольное подмножество Ω , то существует кодирование φ , такое что для произвольного $\theta \in \tilde{\Omega}$ выполняется неравенство:

$$r(N, \theta, \varphi) \leq \frac{h_{1/\sqrt{N}}(\tilde{\Omega})}{n} + \frac{c(\theta)}{n}, \quad (7)$$

где $h_\varepsilon(\tilde{\Omega})$ – ε -энтропия множества источников $\tilde{\Omega}$, которая формально определяется ниже.

В частности, при $\tilde{\Omega} = \Omega$ из (7) вытекает неравенство:

$$r(N, \theta, \varphi) \leq \frac{1}{2} \cdot \frac{\log N}{N} + \frac{c(\theta)}{N}. \quad (8)$$

Если $\beta > \delta > 0$, то в (7) и (8) постоянные не зависят от θ .

3. Энтропия геометрического источника. Кодирование близких источников

Геометрический источник θ_β определяется бесконечно убывающей геометрической прогрессией со знаменателем β , $0 < \beta < 1$, т.е. вероятности появления букв из входного алфавита A равны $P_{\theta_\beta}(a_j) = (1-\beta)\beta^j$, $j = 0, 1, 2, \dots$. Таким образом, каждый из рассматриваемых нами источников однозначно определяется числом из промежутка $(0,1)$.

$$H(\theta_\beta) = - \sum_{j=0}^{\infty} (1-\beta)\beta^j \cdot \log(1-\beta)\beta^j. \quad (9)$$

Найдём $H(\theta_\beta)$, используя (9). Очевидные преобразования позволяют записать $H(\theta_\beta)$ в виде

$$H(\theta_\beta) = - \sum_{j=0}^{\infty} (1-\beta)\beta^j \log(1-\beta) - \sum_{j=0}^{\infty} (1-\beta)\beta^j j \log \beta. \quad (10)$$

Так как $\sum_{j=0}^{\infty} (1-\beta)\beta^j = 1$, то первое слагаемое в (10) равно $-\log(1-\beta)$. Вычислим второе слагаемое из (10)

$$- \sum_{j=0}^{\infty} (1-\beta)\beta^j \cdot j \log \beta = -(1-\beta) \cdot \log \beta \cdot \sum_{j=1}^{\infty} j \beta^j,$$

$$\sum_{j=1}^{\infty} j \cdot \beta^j = \frac{\beta}{(1-\beta)^2}.$$

Получаем:

$$-\sum_{j=0}^{\infty} (1-\beta)\beta^j j \log \beta = -\frac{\beta \cdot \log \beta}{1-\beta}.$$

С учётом вышесказанного из (10) окончательно следует, что

$$H(\theta_\beta) = \frac{H(\beta)}{1-\beta}, \quad (11)$$

где $H(\beta) = -\beta \log \beta - (1-\beta) \log(1-\beta)$.

Расстояние $\rho(\theta_\beta, \theta_{\beta_1})$ между источниками θ_β и θ_{β_1} определим по формуле $\rho(\theta_\beta, \theta_{\beta_1}) = |\beta - \beta_1|$, все аксиомы расстояния выполняются.

Для получения основного результата настоящей работы докажем утверждение, которое позволит нам оценить избыточность кодирования для источников θ_β и θ_{β_Δ} в том случае, когда расстояние между β и β_Δ достаточно мало. Обозначим через φ_β кодирование, которое каждому u , $u \in A^N$, ставит в соответствие слово $\varphi_\beta(u)$ длины

$$|\varphi_\beta(u)| = -\log P_{\theta_\beta}(u), \quad u \in A^N, \quad (12)$$

x – наименьшее целое число, большее или равное x .

Существование дешифруемого кодирования с длинами кодовых слов, удовлетворяющих равенствам (12), гарантирует выполнение для чисел $|\varphi_\beta(u)| = -\log P_\theta(u)$, $u \in A^N$ неравенства Крафта [1], так как

$$\sum_{u \in A^N} 2^{-|\varphi_\beta(u)|} \leq \sum_{u \in A^N} 2^{\log P_{\theta_\beta}(u)} = 1.$$

Сформулируем и докажем основное утверждение настоящего параграфа.

Лемма. Пусть θ_β и θ_{β_1} – произвольные источники, $\beta, \beta_1 \in (0,1)$. Тогда для стоимости $c(N, \varphi_\beta, \theta_{\beta_1})$ кодирования φ_β источника θ_{β_1} при кодировании слов длины N имеет место асимптотическое равенство

$$c(N, \varphi_\beta, \theta_{\beta_1}) \cong \frac{-\beta_1 \log \beta - (1-\beta_1) \log(1-\beta)}{1-\beta_1} + \frac{1}{N}. \quad (13)$$

В частности, при $\beta = \beta_1$ равенство (13) принимает вид:

$$c(N, \varphi_\beta, \theta_\beta) = H(\theta_\beta) + \frac{1}{N}.$$

Доказательство. По определению $c(N, \varphi_\beta, \theta_{\beta_1})$ имеем:

$$c(N, \varphi_\beta, \theta_{\beta_1}) = \frac{1}{N} \sum_{u \in A^N} P_{\theta_{\beta_1}}(u) \cdot -\log P_{\theta_\beta}(u) = \frac{1}{N} \sum_{u \in A^N} P_{\theta_{\beta_1}}(u) \cdot \left(-\log P_{\theta_\beta}(u) \right) + \frac{\lambda}{N}. \quad (14)$$

Здесь λ не зависит от источника и не превосходит 1. Из равенств

$$P_{\theta_\beta}(u) = \prod_{j=0}^{\infty} \left[(1-\beta)\beta^j \right]^{r_j(u)}, \quad \sum_{u \in A^N} P_{\theta_\beta}(u) \cdot r_j(u) = N \cdot (1-\beta)\beta^j$$

и (14) получаем

$$c(N, \varphi_\beta, \theta_{\beta_1}) = \frac{-\beta_1 \log \beta - (1 - \beta_1) \log (1 - \beta)}{1 - \beta_1} + \frac{c(\theta)}{n}.$$

Лемма доказана. □

Перейдем к доказательству основных утверждений настоящей работы.

4. Метод кодирования и оценка его эффективности

Для универсального кодирования произвольного подмножества геометрических источников предложен метод кодирования, основанный на идее, описанной в [9, 10], и лемме предыдущего раздела

В нашем случае предлагаемый метод кодирования состоит в следующем.

Как отмечалось выше, существует взаимно однозначное соответствие между множеством источников и множеством точек интервала $(0,1)$.

Обозначим через $\tilde{\Omega}$ произвольное подмножество геометрических источников. Отсюда следует, что $\tilde{\Omega}$ можно считать подмножеством интервала $(0,1)$.

Пусть θ_β и θ_{β_1} – два произвольных геометрических источника из $\tilde{\Omega}$. Расстояние $\rho(\theta_\beta, \theta_{\beta_1})$ между источниками θ_β и θ_{β_1} определяется равенством

$$\rho(\theta_\beta, \theta_{\beta_1}) = |\beta_1 - \beta|.$$

Обозначим через $h_\varepsilon(\tilde{\Omega})$ ε -энтропию множества $\tilde{\Omega}$. По определению величина $h_\varepsilon(\tilde{\Omega})$ равна логарифму числа элементов минимальной ε -сети для множества источников $\tilde{\Omega}$ [12].

В нашем случае:

а) если Ω – множество всех геометрических источников, то при $\varepsilon \rightarrow 0$

$$h_\varepsilon(\Omega) = \log \frac{1}{\varepsilon} + o\left(\log \frac{1}{\varepsilon}\right);$$

б) если Ω – конечное множество геометрических источников и $\|\Omega\|$ – мощность множества Ω , то при $\varepsilon \rightarrow 0$

$$h_\varepsilon(\Omega) = \log \|\Omega\|.$$

Элементы ε -сети множества источников Ω обозначим $\Omega(\varepsilon)$. По определению $\Omega(\varepsilon)$ для любого источника θ , $\theta \in \Omega$ существует источник θ_1 , $\theta_1 \in \Omega(\varepsilon)$, такой что выполняется неравенство $\rho(\theta, \theta_1) < \varepsilon$. Если источник θ задается числом β , а θ_1 – числом β_1 , то $|\beta - \beta_1| < \varepsilon$.

Докажем теорему, которая в качестве следствия из неё позволит получить оценки избыточности для различных множеств геометрических источников.

Теорема. Для произвольного источника θ , $\theta \in \tilde{\Omega}$, произвольного сколь угодно малого ε , $\varepsilon > 0$, при $N \rightarrow \infty$ существует последовательность кодирований φ_N , такая что для избыточности $r(N, \theta, \varphi_N)$ выполняется асимптотическое неравенство:

$$r(N, \theta, \varphi_N) \lesssim \frac{h_\varepsilon(\tilde{\Omega})}{N} + c_1 \frac{\varepsilon^2}{\beta} + \frac{c_2(\beta)}{N}.$$

Доказательство. Зафиксируем произвольное достаточно малое положительное ε . Пусть $\tilde{\Omega}(\varepsilon)$ – минимальная ε -сеть для $\tilde{\Omega}$. Рассмотрим кодирование φ_N , которое каждому слову u , $u \in A^N$ ставит в соответствие слово $\varphi_N(u)$ длины

$$|\varphi_N(u)| = -\log \frac{\sum_{\tilde{\theta} \in \tilde{\Omega}(\varepsilon)} P_{\tilde{\theta}}(u)}{|\tilde{\Omega}(\varepsilon)|}. \quad (15)$$

Из определения $\tilde{\Omega}(\varepsilon)$ для числа $|\varphi_N(u)|$, $u \in A^N$ выполняется неравенство Крафта [1], следовательно, дешифруемый код с длинами кодовых слов, удовлетворяющих равенствам (15), существует. Пусть $\tilde{\theta}_0$ – фиксированный источник из $\tilde{\Omega}(\varepsilon)$. Используя определение ε -энтропии (14), из (15) несложно получить:

$$|\varphi_N(u)| \leq h_\varepsilon(\tilde{\Omega}) - \log P_{\tilde{\theta}_0}(u) - \log \left(1 + \sum_{\substack{\tilde{\theta} \in \tilde{\Omega}(\varepsilon) \\ \tilde{\theta} \neq \tilde{\theta}_0}} \frac{P_{\tilde{\theta}}(u)}{P_{\tilde{\theta}_0}(u)} \right) + 1.$$

Так как всегда выполнено соотношение

$$-\log \left(1 + \sum_{\substack{\tilde{\theta} \in \tilde{\Omega}(\varepsilon) \\ \tilde{\theta} \neq \tilde{\theta}_0}} \frac{P_{\tilde{\theta}}(u)}{P_{\tilde{\theta}_0}(u)} \right) < 0,$$

то предыдущее неравенство можно переписать в виде:

$$|\varphi_N(u)| \leq h_\varepsilon(\tilde{\Omega}) - \log P_{\tilde{\theta}_0}(u) + 1. \quad (16)$$

Для произвольного источника $\theta_\beta \in \tilde{\Omega}$ существует источник $\tilde{\theta}_0 \in \tilde{\Omega}(\varepsilon)$, такой что расстояние между $\tilde{\theta}_{\beta_0}$ и θ_β не превосходит ε . Из определения $r(N, \theta, \varphi_N)$ и неравенств (16) получаем:

$$r(N, \theta, \varphi_N) \leq \frac{h_\varepsilon(\tilde{\Omega})}{N} - \frac{1}{N} \sum_{u \in A^N} P_{\theta_\beta}(u) \log P_{\tilde{\theta}_{\beta_0}}(u) - H(\theta) + \frac{1}{N}. \quad (17)$$

Так как источник θ_β задается числом β , а источник $\tilde{\theta}_0$ – числом β_0 , то, используя определение стоимости кодирования и лемму, получаем:

$$c(N, \varphi_{\beta_0}, \theta_\beta) = -\frac{1}{N} \sum_{u \in A^N} P_\theta(u) \log P_{\tilde{\theta}_0}(u) = \frac{-\beta \log \beta_0 - (1-\beta) \log(1-\beta_0)}{1-\beta} + \frac{1}{N}. \quad (18)$$

В [9] доказано равенство

$$-\beta \log \beta_0 - (1-\beta) \log(1-\beta_0) \leq -\beta \log \beta - (1-\beta) \log(1-\beta) + \frac{\varepsilon^2}{\beta} + c(\beta) \frac{1}{N^{3/2}}. \quad (19)$$

С учётом (19) соотношение (18) примет вид:

$$c(N, \varphi_{\beta_0}, \theta_\beta) \leq \frac{H(\beta)}{1-\beta} + \frac{\varepsilon^2}{2\beta(1-\beta)} + \frac{c(\beta)}{N^{3/2}}. \quad (20)$$

Согласно (11) первое слагаемое в правой части неравенства (20) равно $H(\theta)$. Отсюда из (17) получаем:

$$r(N, \theta, \varphi_N) \leq \frac{h_\varepsilon(\tilde{\Omega})}{N} + c_1 \frac{\varepsilon^2}{\beta(1-\beta)} + \frac{c_2(\beta)}{N^{3/2}}.$$

Теорема доказана. \square

5. Оценки избыточности универсального кодирования геометрических источников

Теорема, доказанная выше, позволяет получить оценки избыточности универсального кодирования для различных классов геометрических источников. Эта же теорема позволяет построить реальные коды. Сформулированные ниже утверждения являются её следствиями.

Следствие 1. Для множества всех геометрических источников Ω существует последовательность слабоуниверсальных кодирований φ_N , $N = 1, 2, \dots$, такая что для любого источника $\theta \in \Omega$, определяемого знаменателем β , выполняется неравенство:

$$r(N, \theta, \varphi_N) \leq \frac{1}{2} \frac{\log N}{N} + c_1 \frac{1}{\beta N} + \frac{c_2}{N^{3/2}}.$$

Доказательство. Утверждение следствия вытекает из теоремы предыдущего параграфа.

В самом деле, возьмем $\varepsilon = \frac{1}{\sqrt{N}}$, тогда $\varepsilon^2 = \frac{1}{N}$ и $h_{\frac{1}{\sqrt{N}}}(\Omega) = \log N$. Из теоремы получаем

$$r(N, \theta, \varphi_N) \leq \frac{1}{2} \frac{\log N}{N} + c_1 \frac{1}{\beta N}. \quad (21)$$

Из (21) вытекает, что для любого фиксированного источника предложенное кодирование является эффективным, его избыточность стремится к нулю, и имеем скорость убывания $\frac{1}{2} \cdot \frac{\log N}{N}$. Следует отметить, как видно из (21), это стремление к нулю не является равномерным. Следствие доказано. \square

Следующее следствие позволит нам описать множество геометрических источников, для которых существует универсальное кодирование. Обозначим через Ω_δ подмножество геометрических источников θ , для которых знаменатель прогрессии $\beta \geq \delta$. Для множества этих источников справедливо:

Следствие 2. Для избыточности $R(N, \Omega_\delta)$ универсального кодирования множества геометрических источников Ω_δ выполняется неравенство

$$R(N, \Omega_\delta) \leq \frac{1}{2} \frac{\log N}{N} + \frac{c}{N}.$$

c – не зависит от кодируемого источника.

Доказательство. Возьмем произвольный источник θ , $\theta \in \Omega_\delta$. Поэтому знаменатель прогрессии β , определяющий источник, не меньше δ . Из утверждения следствия 1 получаем, что для любого $\theta \in \Omega_\delta$ выполнено неравенство:

$$r(N, \theta, \varphi_N) \leq \frac{1}{2} \frac{\log N}{N} + c_1 \frac{1}{\beta N} + \frac{c_2}{N^{3/2}}.$$

Отсюда и определяется избыточность универсального кодирования $R(N, \Omega)$. Из (5) и (6) имеем

$$R(N, \Omega_\delta) \leq \frac{1}{2} \frac{\log N}{N} + \frac{c}{N},$$

c не зависит от источника.

Следствие доказано. \square

Следствие 3. Для любого множества источников $\tilde{\Omega}$, $\tilde{\Omega} \subset \Omega$ существует слабоуниверсальное кодирование, для которого избыточность $r(N, \theta, \varphi)$ удовлетворяет неравенству

$$r(N, \theta, \varphi) \leq \frac{h_{1/\sqrt{N}}(\tilde{\Omega})}{N} + \frac{c}{\beta N}.$$

Доказательство. Доказательство следует из теоремы при $\varepsilon = \frac{1}{\sqrt{N}}$. Чем ближе β к нулю,

тем большую значимость приобретает второе слагаемое. Однако при фиксированном источнике θ и $N \rightarrow \infty$ для каждого источника θ асимптотически главную роль играет первое слагаемое, т.е. при $N \rightarrow \infty$ всегда выполняется равенство

$$\lim_{N \rightarrow \infty} \frac{c}{\beta N} : \frac{1}{2} \frac{\log N}{N} = \lim_{N \rightarrow \infty} \frac{2}{\beta N} = 0.$$

Следствие доказано. □

Следствие 4. Для избыточности $R(N, \tilde{\Omega}_\delta)$ универсального кодирования множества источников $\tilde{\Omega}_\delta, \tilde{\Omega}_\delta \subset \Omega_\delta$ выполняется неравенство

$$R(N, \Omega_\delta) \leq \frac{h_{1/\sqrt{N}}(\Omega_\delta)}{N} + \frac{c}{N}.$$

Доказательство. Справедливость утверждения непосредственно вытекает из определения $R(N, \Omega_\delta)$ и утверждения теоремы при $\varepsilon = \frac{1}{\sqrt{N}}$.

Следствие доказано. □

6. Заключение

В данной работе было рассмотрено множество геометрических источников с бесконечным алфавитом. Доказано, что для произвольного множества этих источников существует слабо универсальное кодирование. Для множества источников со знаменателем не меньше чем δ , $\delta > 0$ доказано существование универсального кодирования. Получены оценки убывания избыточности в зависимости от ε -энтропии множества геометрических источников.

Литература

1. Левенштейн В. И. Об избыточности и замедлении разделимого кодирования натуральных чисел // Проблемы кибернетики. 1968. В. 20. С. 173–179.
2. Gyorfi L., Pali I.; van der Meulen E. C. There is no Universal Source Code for an Infinite Source Alphabet // IEEE Transactions on Information Theory. 1994. V. 40, № 1. P. 267–271.
3. Devison L. D. Universal Noiseless coding // IEEE Transactions on Information Theory. 1973. V. 19, № 6. P. 783–795.
4. Shtarkov Yu. M., Babkin V. E. Combinatorial encoding for discrete stationary sources // Proc. 2nd Inter. Symp. on Inform. Theory, Tsahkad-zor, Armenia, USSR, 1971. P. 249–256.
5. Бабкин В. Ф. Универсальное кодирование дискретных независимых источников со счетным алфавитом // Вопросы кибернетики. 1977. В. 34. С. 3–10.
6. Кудряшов В. Д., Поров А. В. Универсальное кодирование источников без памяти со счетным алфавитом // Проблемы передачи информации. 2014. Т. 50, № 4. С. 100–108.
7. Штарков Ю. М., Рыбьева Н. Е. Адаптивное кодирование для пуассоновских источников // Кодирование в сложных системах. 1974. С. 169–175.
8. Трофимов В. К. Кодирование потоков Пуассона при неизвестных и неточно известных параметрах // Проблемы передачи информации. 1987. Т. 23, № 4. С. 49–57.
9. Трофимов В. К. Универсальное кодирование произвольного множества источников без памяти // Вестник СибГУТИ. 2018. № 4. С. 30–34.
10. Трофимов В. К. Оценка избыточности универсального кодирования произвольного множества источников без памяти // Вестник СибГУТИ. 2019. № 4. С. 56–61.

11. *Файнштейн А.* Основы теории информации. М., 1960. 138 с.
12. *Витушкин А. Г.* Оценка сложности задачи табулирования М.: Гос. изд. физ.-мат. лит., 1959. 228 с.

*Статья поступила в редакцию 17.02.2021;
переработанный вариант – 01.06.2021.*

Трофимов Виктор Куприянович

д.т.н., профессор, декан факультета информатики и вычислительной техники, заведующий кафедрой высшей математики СибГУТИ (630102, Новосибирск, ул. Кирова, 86), тел. (383) 269-82-70; e-mail: trofimov@sibguti.ru.

Область научных исследований – теория информации, кодирование информации, порождаемой дискретным источником, оценка сложности кодирования.

Encoding geometric sources with unknown statistics

V. K. Trofimov

Universal encoding method of an arbitrary set of sources without memory generating letters of an infinite alphabet is proposed. The probabilities of the input alphabet letter appearance are a geometric progression. The proposed method is weakly universal for the set of all geometric sources. If the denominator of the geometric progression exceeds δ , $\delta > 0$, the proposed encoding is universal. Redundancy estimates are obtained for an arbitrary subset of geometric sources.

Keywords: encoding, redundancy, entropy, message source.