УДК 621.391.15

ОПТИМАЛЬНОЕ РАВНОМЕРНОЕ ПО ВЫХОДУ КОДИРОВАНИЕ ДЛЯ ОБЪЕДИНЕНИЯ РАЗЛИЧНЫХ МНОЖЕСТВ ИСТОЧНИКОВ

В. К. Трофимов[1,2], Т. В. Храмова[1]

[1] *Сибирский государственный университет телекоммуникаций и информатики, 630102, г. Новосибирск, ул. Кирова, 86*

[2] *Институт систем информатики им. А. П. Ершова СО РАН, 630090, г. Новосибирск, просп. Академика Лаврентьева, 6*

E-mail: trofimov@sibsutis.ru
tvkhramova@gmail.com

Предложен метод оптимального равномерного по выходу кодирования для множества источников, состоящего из объединения счётного числа множеств источников. Показано, что кодирование для объединения марковских источников с конечной памятью является асимптотически оптимальным. Установлено, что рассматриваемое кодирование — слабоуниверсальное для множества стационарных источников.

Ключевые слова: кодирование, избыточность, энтропия, хранение и обработка информации, источник сообщений.

DOI: 10.15372/AUT20170107

Введение. Проблемы сжатия (кодирования) информации [1] относятся к фундаментальным в области инфокоммуникаций. В [2] отмечено, что решение этих проблем значимо при создании большемасштабных распределённых вычислительных систем. В таких системах для сжатия информации, как правило, используют информационно-вычислительные технологии.

В предлагаемой работе исследуется вопрос об универсальном равномерном по выходу кодировании для объединения различных множеств источников. Для универсального равномерного по входу, но неравномерного по выходу кодирования данная задача впервые решена в [3]. Кодирование символами различной стоимости было названо дважды универсальным и рассматривалось в [4].

Равномерное по выходу кодирование имеет ряд преимуществ в сравнении с неравномерным. В частности, в этом случае отсутствует ошибка синхронизации и кодирование удобно для дальнейшего применения корректирующих кодов.

Равномерное по выходу кодирование при известной статистике сообщений изучалось в [5–9], а при неизвестной — в [10–18].

Цель данного исследования — разработка метода универсального равномерного по выходу кодирования и доказательство его оптимальности для источников, принадлежащих различным классам. Метод оптимален для бернуллиевских и для марковских источников любой фиксированной связности, при этом кодирующее и декодирующее устройства настраиваются 1 раз и не требуют перенастройки. Также отметим, что предложенная универсальность не усложняет процессы кодирования и декодирования.

Основные определения и обозначения. Пусть буквы конечного алфавита $A = \{a_1, a_2, \ldots, a_k\}$, $2 \le k < \infty$, порождаются источником θ. Мера, заданная на последовательности букв, определяет его тип. Если вероятности порождения букв независимы, то источник называют бернуллиевским. В этом случае $P_\theta(a_j) = \theta_j$, $\theta_1 + \theta_2 + \ldots + \theta_k = 1$. Если вероятность появления очередной буквы зависит от предыдущей, т. е. $P_\theta(a_i/a_j) = \theta_{ji}$, $\sum_{i=1}^{k} \theta_{ji} = 1$, то источник называют марковским. Если вероятность появления очередной буквы зависит от s предшествующих букв, т. е. $P_\theta(a_i/v) = \theta_{vi}$, где $v \in A^s$, то источник θ является марковским с памятью s. Следует отметить, что для любого слова $v \in A^s$, $0 \le s < \infty$, выполняется равенство $\sum_{i=1}^{k} \theta_{vi} = 1$. Множество всех марковских источников с памятью s обозначим Ω_s.

Пусть u — произвольное слово в алфавите A, тогда $P_\theta(u)$ — вероятность слова u, порождённого источником θ. Число букв в слове u назовём его длиной и обозначим $|u|$. Пусть $H(\theta)$ — энтропия источника θ. Как известно [19, 20], если θ — стационарный источник, то

$$H(\theta) = -\lim_{n \to \infty} \frac{1}{n} \sum_{u \in A^n} P_\theta(u) \log P_\theta(u). \tag{1}$$

Здесь и далее $\log x = \log_2 x$, $0 \log 0 = 0$.

Для бернуллиевского источника θ из (1) следует, что его энтропия, обозначаемая далее $H_0(\theta)$, вычисляется по формуле

$$H_0(\theta) = -\sum_{i=1}^{k} \theta_i \log \theta_i. \tag{2}$$

Если θ — марковский источник с памятью s, то его энтропию обозначим $H_s(\theta)$. Имеет место равенство

$$H_s(\theta) = -\sum_{v \in A^s} \theta_{0v} \sum_{i=1}^{k} \theta_{vi} \log \theta_{vi}, \tag{3}$$

где θ_{0v} — начальные стационарные вероятности слов v, $v \in A^s$. При $s = 0$ из (3) получим соотношение (2).

Пусть T — множество слов во входном алфавите. Множество T полное, если оно префиксное, и при любом непустом слове u (в алфавите A) множество слов $T \cup \{u\}$ не префиксное. Конечное полное множество T назовём кодовым. Например, множество всех слов длины n, взятых в алфавите A, т. е. A^n — кодовое, а множество $A^n \backslash \{\underbrace{a_1, \dots, a_1}_{n}\}$ не является кодовым, потому что оно неполное.

Пусть θ — произвольный источник из Ω_s. Обозначим через $\theta(T)$ марковскую цепь, состояниями которой служат слова из T, а переходные вероятности $P_{\theta(T)}(u/v)$, $u, v \in$ $\in T$, индуцируются источником θ. Будем рассматривать только марковские источники с памятью s, переходные вероятности которых строго положительны. В этом случае для любых $u, v \in T$ выполняются неравенства $P_{\theta(T)}(u/v) > 0$, поэтому для марковской цепи $\theta(T)$ существует стационарное распределение $P^0_{\theta(T)}(u) > 0$, $u \in T$. Средняя длина слова $d_s(T, \theta)$ для множества T, как доказано в [21], вычисляется по формуле

$$d_s(T, \theta) = \sum_{u \in T} P^0_{\theta(T)}(u) |u|. \tag{4}$$

Полубесконечная последовательность букв, порождаемая источником θ, однозначно разбивается на последовательность слов из фиксированного кодового множества T. Полученная последовательность слов из T с помощью отображения φ переводится в слова выходного алфавита B, который, не уменьшая общности, можно считать двоичным. Из неравенства Макмиллана — Крафта [19, 20] следует, что самое общее из всех возможных дешифруемых кодирований φ такое, в котором множество слов в выходном алфавите $\varphi(T) = \{\varphi(u),\ u \in T\}$ является префиксным. Если длины всех слов некоторого множества D равны между собой, то считается, что D состоит из блоков, в противном случае — из слов переменной длины. В зависимости от видов множества T и $\varphi(T)$ логически возможны следующие виды кодирований:

1) отображающее блоки в слова переменной длины (обозначается bv);
2) отображающее слова переменной длины в блоки (vb);
3) отображающее слова переменной длины в слова переменной длины (vv);
4) отображающее блоки в блоки (bb).

Итак, всякое кодирование φ однозначно определяется тройкой $T, \varphi, \varphi(T)$. Среднее число букв выходного алфавита при кодировании типа σ ($\sigma = bv, vb, vv$), приходящихся на одну букву входного, назовём стоимостью кодирования и обозначим через $C_\sigma(T, \theta, \varphi)$. В [21] доказано, что стоимость кодирования типа σ ($\sigma = bv, vb, vv$) для произвольного кодового множества T и любого источника θ, $\theta \in \Omega_s$, $0 \le s < \infty$, находится из равенства

$$C_\sigma(T, \theta, \varphi) = \frac{1}{d_s(T, \theta) - \hat{s} + 1} \sum_{u \in T} P^0_{\theta(T)}(u)|\varphi(u)|, \tag{5}$$

где величина $d_s(T, \theta)$ задаётся формулой (4); $|\varphi(u)|$ — число букв в слове $\varphi(u)$.

Эффективность кодирования φ, как обычно [1, 2, 6, 10, 12, 19, 20], будем оценивать разностью между стоимостью кодирования $C_\sigma(T, \theta, \varphi)$ и энтропией источника $H(\theta)$. Эта разность называется избыточностью кодирования и обозначается $r_\sigma(T, \theta, \varphi)$, т. е.

$$r_\sigma(T, \theta, \varphi) = C_\sigma(T, \theta, \varphi) - H(\theta). \tag{6}$$

Величину

$$R_\sigma(N, \Omega) = \sup_{\theta \in \Omega} \inf_{\varphi} r_\sigma(T, \theta, \varphi), \tag{7}$$

где нижняя грань берётся по всем кодированиям φ, для которых кодовое множество T имеет не более чем k^N слов, назовём избыточностью универсального кодирования типа σ для множества источников Ω сложности N. Построение оптимального кодирования при заданной сложности — основной вопрос при изучении передачи сообщений по каналу без шума. Решение поставленной задачи позволяет ответить на вопрос, какой избыточности можно достигнуть при заданной сложности кодирования.

Если множество источников $\{\theta\}$ состоит из единственного источника, то это кодирование известного источника, эффективность которого согласно (7) определяется величиной $R_\sigma(N, \theta)$. Поведение $R_\sigma(N, \theta)$ подробно изучено для всех кодирований, например, в работах [1, 6–9]. Универсальное кодирование марковских источников различных типов также хорошо исследовано [10–19, 22].

Полагаем, что кодирование φ типа σ является универсальным для множества источников Ω, если $R_\sigma(N, \theta)$ стремится к нулю равномерно относительно Ω. Если же $R_\sigma(N, \theta)$ сходится к нулю при любом $\theta \in \Omega$, но при этом нет равномерной сходимости, то будем считать, что на множестве источников Ω существует слабоуниверсальное кодирование.

Сосредоточим своё внимание на универсальном равномерном по выходу кодировании для множества источников Ω, которое является объединением марковских источников с конечной памятью:

$$\Omega = \bigcup_{s=0}^{\infty} \Omega_s. \tag{8}$$

Предложенное в данной работе кодирование при выборе источника из Ω_s, $0 \le s < \infty$, обладает всеми достоинствами кода из [18], которые сохраняются при любом выборе s, $0 \le s < \infty$. Кроме того, оно является слабоуниверсальным для множества всех стационарных источников [16, 22].

Покажем, что существует последовательность vb кодирований φ_N с областью определения T_N, с числом кодовых слов $\|T_N\|$, $\|T_N\| \le k^N$, для которых избыточность $r_{vb}(T_N, \theta, \varphi_N)$ при любом источнике $\theta \in \Omega$ стремится к нулю. При этом, если $\theta \in \Omega_s$, скорость убывания избыточности $r_{vb}(T_N, \theta, \varphi_N)$ совпадает со скоростью убывания $R_{vb}(T_N, \Omega_s)$. Другими словами, предложенное кодирование автоматически перестраивает кодирующее устройство на оптимальный вариант работы. Платой за эту перестройку служит аддитивная величина, пропорциональная $\lambda / \log \|T_N\|$, λ — постоянная, зависящая только от s.

Как было отмечено выше, марковский источник θ связанности s задаётся начальным распределением вероятностей θ_{0v} появления блока v за первые s шагов работы источника и вероятностями θ_{vi} появления буквы a_i после блока v, $a_i \in A$, $v \in A^s$. Отсюда следует, что вероятность $P_\theta(u)$ порождения источником θ слова u, $|u| > s$, начинающегося блоком v, $v \in A^s$, определяется равенством

$$P_\theta(u) = \theta_{0v} \prod_{v \in A^s} \prod_{i=1}^{k} \theta_{vi}^{r_{vi}(u)}. \tag{9}$$

В (9) и далее $r_{vi}(u)$ — число вхождений блоков vi в слово u.

На множестве источников Ω_s зададим распределение с плотностью [6]

$$\omega_s(\theta) = \Big(\frac{\Gamma(k/2)}{k\pi^{k/2}}\Big)^{k^s} 1 \Big/ \sqrt{\prod_{v \in A^s} \prod_{i=1}^{k} \theta_{vi}}. \tag{10}$$

Проинтегрировав вероятность слова u, порождённого источником θ, по множеству источников Ω_s, если на Ω_s задана плотность $\omega_s(\theta)$, $s < |u|$, и используя (9), (10), получим [10]

$$\bar{P}_s(u) = \Big[\frac{\Gamma(k/2)}{k\pi^{k/2}}\Big]^{k^s} \prod_{v \in A^s} \frac{\prod_{j=1}^{k} \Gamma(r_{vj}(u) + 1/2)}{\Gamma(r_v(u) + k/2)}, \tag{11}$$

где $\Gamma(z)$ — гамма-функции от z; $r_v(u) = \sum_{j=1}^{k} r_{vj}(u)$.

Величина $\bar{P}_s(u)$, определяемая равенством (11), есть средняя вероятность слова u по множеству источников Ω_s с заданной на нём плотностью распределения $\omega_s(\theta)$ (10).

Введём в рассмотрение среднюю вероятность $\bar{P}(u)$ слова u, вычисляемую из равенства

$$\bar{P}(u) = \frac{6}{\pi^2} \sum_{s=0}^{\infty} \frac{1}{(s+1)^2} \bar{P}_s(u), \tag{12}$$

где $6/\pi^2$ — нормирующий множитель.

Отметим, что если T — произвольное кодовое множество, то выполняются следующие соотношения:

$$\sum_{u \in T} \bar{P}(u) = 1; \qquad \lim_{|u| \to \infty} \bar{P}(u) = 0. \tag{13}$$

Основное утверждение и его доказательство. Сформулируем и докажем следующее утверждение.

Теорема 1. Для множества источников $\Omega = \bigcup\limits_{s=0}^{\infty} \Omega_s$ существует такая последовательность универсальных равномерных по выходу кодирований φ_N, что для любого источника $\theta \in \Omega_s$ избыточность кодирования $r_{vb}(T_N, \theta, \varphi_N)$ удовлетворяет неравенству при $N \to \infty$:

$$r_{vb}(T_N, \theta, \varphi_N) \le \frac{k^s(k-1)+2}{2}\, H(\theta) \frac{\log\log \|T_N\|}{\log \|T_N\|}\,(1+o(1)) + \frac{2\log s}{\log \|T_N\|}$$

($\|T_N\|$ — мощность множества T_N).

Доказательство. Как уже отмечалось, каждое кодирование задаётся тройкой $T, \varphi, \varphi(T)$, где T — область определения, $\varphi(T)$ — множество значений отображения φ. Для равномерного по выходу кодирования множество значений отображения φ находится достаточно просто. В этом случае $\varphi(T) = B^{\lceil \log \|T\| \rceil}$, где $\lceil x \rceil$ — наименьшее целое, большее или равное x. Таким образом, при построении равномерных по выходу кодирований вся сложность заключается в формировании кодовых множеств T_N, $N = 1, 2, \ldots$. Зафиксируем произвольное натуральное N, в кодовое множество T_N включим все слова u, для которых справедливо неравенство

$$1/\bar{P}(u) \le k^N, \tag{14}$$

в то же время существует буква a_j, $a_j \in A$, такая, что для конкатенации слова u и буквы a_j выполняется неравенство

$$1/\bar{P}(ua_j) > k^N. \tag{15}$$

В силу второго соотношения из (13) описанный процесс конечен.

Совершенно очевидно, что построенное таким образом множество T_N является конечным полным множеством слов во входном алфавите, т. е. T_N — кодовое множество. При равномерном по выходу кодировании каждому слову u ставится в соответствие слово $\varphi_N(u)$, $u \in T_N$, длиной $\lceil \log \|T_N\| \rceil$, другими словами, при любом $u \in T_N$ выполняется равенство

$$|\varphi_N(u)| = \lceil \log \|T_N\| \rceil. \tag{16}$$

Оценим избыточность предложенного метода кодирования при $N \to \infty$. Из определения избыточности (6) и из (16) при $\theta \in \Omega_s$, $0 \le s \le L < \infty$, имеем

$$r_{vb}(T_N, \theta, \varphi_N) = \frac{\lceil \log \|T_N\| \rceil}{d_s(T_N, \theta) - \hat{s} + 1} - H(\theta). \tag{17}$$

Кодирование φ_N дешифруемое, поэтому величина $r_{vb}(T_N, \theta, \varphi_N)$ неотрицательна. Найдём верхнюю оценку этой величины. Из (14) следует, что при любом u, $u \in T_N$, справедливо неравенство $\bar{P}(u) \geq 1/k^N$, просуммировав которое по всем словам u из T_N, имеем

$$\sum_{u \in T_N} \bar{P}(u) \geq \sum_{u \in T_N} \frac{1}{k^N}.$$

Учитывая первое соотношение (13), из предыдущего соотношения получим

$$k^N \geq \|T_N\|. \tag{18}$$

Принимая в расчёт (18) и (15), из (17) следует

$$r_{vb}(T_N, \theta, \varphi_N) \leq \frac{\sum\limits_{u \in T_N} P^0_{\theta(T_N)}(u) \log \|T_N\|}{d_s(T_N, \theta) - \hat{s} + 1} - H(\theta) + \frac{1}{d_s(T_N, \theta) - \hat{s} + 1} \leq$$

$$\leq \frac{\sum\limits_{u \in T_N} P^0_{\theta(T_N)}(u) \log k^N}{d_s(T_N, \theta) - \hat{s} + 1} - H(\theta) + \frac{1}{d_s(T_N, \theta) - \hat{s} + 1} \leq$$

$$\leq \frac{-\sum\limits_{u \in T_N} P^0_{\theta(T_N)}(u) \log \bar{P}(ua_j)}{d_s(T_N, \theta) - \hat{s} + 1} - H(\theta) + \frac{1}{d_s(T_N, \theta) - \hat{s} + 1}. \tag{19}$$

Учитывая, что средняя вероятность $\bar{P}(ua_j)$ слова ua_j определяется (12), для слова u, $|u| < \min\limits_{v \in T_N} \{|v|, s\}$, справедливо равенство

$$\bar{P}(ua_j) = \sum_{\substack{t=0 \\ t \neq s}}^{L} \frac{6}{\pi^2 (t+1)^2} \bar{P}_t(ua_j) + \frac{6}{\pi^2 (s+1)^2} \bar{P}_s(ua_j).$$

Из вычисления $\bar{P}_s(ua_j)$ (11) имеем $\frac{6}{\pi^2(s+1)^2} \bar{P}_s(ua_j) \neq 0$, поэтому последнее равенство перепишем в виде

$$\bar{P}(ua_j) = \frac{6}{\pi^2 (s+1)^2} \bar{P}_s(ua_j) \Big(1 + \frac{6}{\pi^2}(s+1)^2 \sum_{\substack{t=0 \\ t \neq s}}^{L} \frac{1}{(t+1)^2} \bar{P}_t(ua_j)/\bar{P}_s(ua_j)\Big). \tag{20}$$

Из (19) и (20) вытекает

$$r_{vb}(T_N, \theta, \varphi_N) \leq \frac{-\sum\limits_{u \in T_N} P^0_{\theta(T_N)}(u) \log \bar{P}_s(ua_j)}{d_s(T_N, \theta) - \hat{s} + 1} - H(\theta) + \frac{2 \log(s+1) - \log(6/\pi^2) + 1}{d_s(T_N, \theta) - \hat{s} + 1} -$$

$$- \frac{1}{d_s(T_N, \theta) - \hat{s} + 1} \log \Big(1 + \frac{6}{\pi^2}(s+1)^2 \sum_{\substack{t=0 \\ t \neq s}}^{L} \frac{1}{(t+1)^2} \frac{\bar{P}_t(ua_j)}{\bar{P}_s(ua_j)}\Big). \tag{21}$$

Так как

$$1+\frac{6}{\pi^2}\,(s+1)^2\sum_{\substack{t=0\\ t\neq s}}^{\infty}\frac{1}{(t+1)^2}\,\frac{\bar P_t(ua_j)}{\bar P_s(ua_j)}>1,$$

то

$$-\log\Bigl(1+\frac{6}{\pi^2}\,(s+1)^2\sum_{\substack{t=0\\ t\neq s}}^{\infty}\frac{1}{(t+1)^2}\,\frac{\bar P_t(ua_j)}{\bar P_s(ua_j)}\Bigr)<0.$$

Отсюда и из (21) получаем

$$r_{vb}(T_N,\theta,\varphi_N)\le\frac{-\sum\limits_{u\,\in\,T_N}P^0_{\theta(T_N)}(u)\log\bar P_s(ua_j)}{d_s(T_N,\theta)-\hat s+1}-H(\theta)+\frac{2\log(s+1)-\log(6/\pi^2)+1}{d_s(T_N,\theta)-\hat s+1}.\quad(22)$$

Из определения $\bar P_s(u)$ (11) имеем

$$-\log\bar P_s(ua_j)=-\log\bar P_s(u)+\log\frac{r_v(ua_j)+k/2}{r_{vj}+3/2}.\quad(23)$$

Так как $r_v(ua_j)+k/2\le|u|+k/2+1$, $r_{vj}+3/2\ge1$, равенство (23) запишем в виде неравенства

$$-\log\bar P_s(ua_j)\le-\lceil\log\bar P_s(u)\rceil+\log(|u|+k/2+1).$$

Используя последнее соотношение, из (22) имеем

$$r_{vb}(T_N,\theta,\varphi_N)\le r_{vv}(T_N,\theta,\varphi^s_N)+$$

$$+\frac{-\sum\limits_{u\,\in\,T_N}P^0_{\theta(T_N)}(u)\log(|u|+k/2+1)}{d_s(T_N,\theta)-\hat s+1}+\frac{2\log(s+1)-\log(6/\pi^2)+1}{d_s(T_N,\theta)-\hat s+1}.\quad(24)$$

Здесь $r_{vv}(T_N,\theta,\varphi^s_N)$ — избыточность неравномерного по входу и выходу кодирования источника θ с кодовым множеством T_N и отображением φ^s_N, которое каждому $u\in T_N$ ставит в соответствие слово $\varphi^s_N(u)$ длиной $|\varphi^s_N(u)|=\lceil-\log\varphi^s_N(u)\rceil$. Для этой величины в [23] установлена оценка

$$r_{vv}(T_N,\theta,\varphi^s_N)\le\frac{k^s(k-1)}{2}\,\frac{\log d_s(T_N,\theta)}{d_s(T_N,\theta)}+\frac{C}{d_s(T_N,\theta)},$$

где C не зависит ни от θ, ни от N.

Учитывая это неравенство и применяя неравенство Йенсена для функции $\log x$, из (24) имеем

$$r_{vb}(T_N,\theta,\varphi_N)\le\frac{k^s(k-1)+2}{2}\,\frac{\log d_s(T_N,\theta)}{d_s(T_N,\theta)}+\frac{2\log(s+1)}{d_s(T_N,\theta)}+\frac{C_1}{d_s(T_N,\theta)}.\quad(25)$$

В [18] доказано, что $d_s(T_N,\theta)=\log\|T_N\|/(H(\theta)+\alpha_N(\theta))$, где $\alpha_N(\theta)\to0$ при $N\to\infty$. Отсюда и из (25) запишем

$$r_{vb}(T_N,\theta,\varphi_N)\le\frac{k^s(k-1)+2}{2}\,H(\theta)\frac{\log\log\|T_N\|}{\log\|T_N\|}\,(1+o(1))+\frac{2\log(s+1)}{\log\|T_N\|}\,(H(\theta)+o(1)).$$

Теорема доказана.

Равномерное по выходу кодирование стационарных источников. В работах [16, 18] доказано существование слабоуниверсального равномерного по выходу кодирования для класса всех стационарных источников. Покажем, что построенное выше кодирование слабоуниверсальное для множества всех стационарных источников и обладает рядом преимуществ в сравнении с кодами из [16, 18]. Более точно и справедливо следующее утверждение.

Теорема 2. Предложенное равномерное по выходу кодирование φ_N является слабоуниверсальным для множества всех стационарных источников Ω_∞.

Доказательство. Каждый стационарный источник θ, $\theta \in \Omega_\infty$, задаётся условными вероятностями $P_s(a_i \,|\, v)$, $a_i \in A$, $v \in A^s$, $s = 0, 1, 2, \ldots$, появления буквы a_i после блока v. Таким образом, каждый стационарный источник θ определяет последовательность марковских источников θ_s, $s = 0, 1, 2, \ldots$. При s, стремящемся к бесконечности, энтропия $H(\theta_s)$ источника θ_s, не возрастая, сходится к энтропии $H(\theta)$ источника θ, точнее, справедливы соотношения

$$H(\theta_0) \geq H(\theta_1) \geq H(\theta_2) \geq \ldots \geq H(\theta_s) \geq H(\theta_{s+1}) \geq \ldots, \quad \lim_{s\to\infty} H(\theta_s) = H(\theta). \tag{26}$$

Пусть φ_N — кодирование, построенное выше с областью определения T_N. Тогда для любого источника $\theta \in \Omega_\infty$ фиксированного s, $0 \leq s < \infty$, найдена стоимость кодирования $C_{vb}(T_N, \theta_s, \varphi_N)$ (5). Покажем, что при N и s, стремящихся к бесконечности, стоимость кодирования $C_{vb}(T_N, \theta_s, \varphi_N)$ существует и равна энтропии источника $H(\theta)$. При этом необходимо установить, что для стационарного источника θ, $\theta \in \Omega_\infty$, $r_{vb}(T_N, \theta_s, \varphi_N) \to 0$ с ростом N и s [19, 20].

Пусть $\theta \in \Omega_\infty$ и $\theta_s \in \Omega_s$, $s = 0, 1, 2, \ldots$, — марковские источники, определяющие θ, тогда

$$r_{vb}(T_N, \theta_s, \varphi_N) = \frac{\lceil \log \|T_N\| \rceil}{d_s(T_N, \theta_s) - \hat{s} + 1} - H(\theta),$$

т. е.

$$r_{vb}(T_N, \theta_s, \varphi_N) = \Big[\frac{\lceil \log \|T_N\| \rceil}{d_s(T_N, \theta_s) - \hat{s} + 1} - H(\theta_s)\Big] + [H(\theta_s) - H(\theta)]. \tag{27}$$

В равенстве (27) первое слагаемое в правой части согласно теореме 1 ограничено асимптотически сверху величиной $\frac{k^s(k-1)+2}{2}\,\frac{\log\log\|T_N\|}{\log\|T_N\|}$ и, значит, с ростом N стремится к нулю. Из (26) следует, что с ростом s второе слагаемое в (27) также сходится к нулю. Если выбрать $s = o(\log\|T_N\| - \log\log\|T_N\|)$, то первое и второе слагаемые из правой части равенства (27) стремятся к нулю одновременно с ростом N. Тогда из (27) вытекает

$$\lim_{N\to\infty} r_{vb}(T_N, \theta_s, \varphi_N) = 0,$$

т. е.

$$\lim_{N\to\infty} C_{vb}(T_N, \theta_s, \varphi_N) = H(\theta).$$

Теорема доказана.

Заключение. Предложен метод неравномерного по входу, но равномерного по выходу кодирования информации, порождённой неизвестным источником, принадлежащим объединению различных множеств марковских источников. Такое кодирование является оптимально-универсальным для каждого множества источников, входящих в данное объединение, а также слабоуниверсальным для множества стационарных источников.

СПИСОК ЛИТЕРАТУРЫ

1. **Шеннон К.** Работы по теории информации и кибернетике. М.: Изд-во иностранной литературы, 1963. С. 243–332.
2. **Хорошевский В. Г.** Архитектура вычислительных систем. М.: Изд-во МГТУ им. Н. Э. Баумана, 2008. 520 с.
3. **Рябко Б. Я.** Дважды универсальное кодирование // Проблемы передачи информации. 1984. **20**, вып. 3. С. 24–28.
4. **Трофимов В. К., Храмова Т. В.** Оптимальное универсальное кодирование для объединения различных множеств источников символами неравной длительности // Вестн. СибГУТИ. 2014. № 4. С. 30–36.
5. **Блох Э. Л.** О передаче бинарной последовательности равномерным кодом // Проблемы передачи информации. 1960. Вып. 5. С. 12–22.
6. **Ходак Г. Л.** Оценки избыточности при пословном кодировании сообщений, порождаемых бернуллиевским источником // Проблемы передачи информации. 1972. **8**, вып. 2. С. 21–32.
7. **Jelinek F., Schneider K.** On variable-length-to-block coding // IEEE Trans. Inform. Theory. 1972. **18**, Is. 6. P. 756–774.
8. **Трофимов В. К.** Эффективное кодирование блоками слов различной длины, порожденных известным марковским источником // Обработка информации в системах связи. Л.: ЛЭИС, 1985. Т. 29. С. 9–15.
9. **Ziv J.** Variable-to-fixed length codes are better than fixed-to-variable length codes for Markov sources // IEEE Trans. Inform. Theory. 1990. **36**, Is. 4. P. 861–863.
10. **Krichevsky R. E., Trofimov V. K.** The performance of universal encoding // IEEE Trans. Inform. Theory. 1981. **27**, Is. 2. P. 199–207.
11. **Потапов В. Н.** Обзор методов неискажающего кодирования дискретных источников // Дискретный анализ и исследование операций. 1999. Сер. 1. **6**, № 4. С. 49–91.
12. **Штарьков Ю. М.** Универсальное кодирование. Теория и алгоритмы. М.: Физматлит, 2013. 288 с.
13. **Трофимов В. К.** Универсальное равномерное по выходу кодирование бернуллиевских источников // Методы дискретного анализа в теории кодов и схем. Новосибирск: ИМ СО АН СССР, 1976. Вып. 29. С. 87–99.
14. **Lawrence J. C.** A new universal coding scheme for the binary memoryless source // IEEE Trans. Inform. Theory. 1977. **23**, Is. 4. P. 446–472.
15. **Штарьков Ю. М.** Равномерное по выходу универсальное кодирование дискретных источников без памяти // Проблемы передачи информации. 1991. **27**, вып. 1. С. 3–13.
16. **Трофимов В. К.** Слабоуниверсальное равномерное по выходу кодирование дискретных стационарных источников // Вестн. СибГУТИ. 2010. № 2. С. 101–111.
17. **Трофимов В. К.** Об эффективности равномерного по выходу кодирования бернуллиевских источников при неизвестной статистике сообщений // Автометрия. 2010. **46**, № 6. С. 32–39.
18. **Трофимов В. К.** Об эффективности равномерного по выходу кодирования марковских источников при неизвестной статистике сообщений // Автометрия. 2015. **51**, № 3. С. 11–18.

19. **Фано Р.** Передача информации. Статистическая теория связи. М.: Мир, 1965. 440 с.
20. **Галлагер Р.** Теория информации и надежная связь. М.: Сов. радио, 1974. 720 с.
21. **Могульский А. А., Трофимов В. К.** Тождество Вальда и стоимость кодирования для цепей Маркова // Докл. VII Всесоюз. конф. по теории кодирования и передачи информации. Москва — Вильнюс, 1978. Ч. I. Теория информации. С. 112–116.
22. **Трофимов В. К.** Равномерное по выходу кодирование дискретных стационарных источников сообщений с неизвестной статистикой // Вестн. ТГУ. Управление, вычислительная техника и информатика. 2011. № 1(14). С. 55–63.
23. **Трофимов В. К., Агульник В. И., Резван И. И.** Пословное кодирование сообщений, порождённых стационарным источником с неизвестной статистикой // Ползуновский вестн. 2012. № 2/1. С. 115–121.

Поступила в редакцию 10 июня 2016 г.