

Слабоуниверсальное равномерное по выходу кодирование дискретных стационарных источников^{*)}

В.К. Трофимов

Доказано существование слабоуниверсального равномерного по выходу кодирования для множества всех стационарных дискретных источников. Получены необходимые и достаточные условия существования универсального равномерного по выходу кодирования множества стационарных дискретных источников.

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

1. Основные определения. Постановка задачи

Из-за возрастающего потока и объёма информации большое значение имеет проблема сжатия (кодирования) информации [1]. Решение этих проблем значимо и при создании большемасштабных распределённых вычислительных систем [2]. Методы сжатия информации в таких системах, как правило, используют параллельные информационно-вычислительные технологии.

Настоящая работа посвящена кодированию информации, порождённой источником, в её классической форме, предложенной К. Шенноном [1]. Для постановки задачи и формулировки основных утверждений приведём основные определения и обозначения.

Пусть буквы конечного алфавита $A = \{a_1, a_2, \dots, a_k\}$, $2 \leq k < \infty$, порождаются источником θ . Мера, заданная на последовательности букв, порождаемой источником, определяет тип источника. Если вероятности порождения букв независимы, то источник называют бернуллиевским. В этом случае $P_\theta(a_j) = \theta_j$, $\theta_1 + \theta_2 + \dots + \theta_k = 1$. Если же вероятность появления очередной буквы зависит от предыдущей, то $P_\theta(a_i/a_j) = \theta_{ij}$, $\sum_{i=1}^k \theta_{ij} = 1$, $i = \overline{1, k}$, и в этом случае источник называют марковским. Если вероятность появления очередной буквы зависит от s предшествующих букв, т.е. $P_\theta(a_j/v) = \theta_{vj}$, где $v \in A^s$, то источник θ называют марковским с памятью s . Следует отметить, что для любого слова $v \in A^s$, $0 \leq s < \infty$, выполняется равенство $\sum_{j=1}^k \theta_{vj} = 1$. Множество всех марковских источников с памятью s обозначим Ω_s . Дискретный стационарный источник θ задаётся всеми условными распределениями вероятностей $P_\theta(a_j/v) = \theta_{vj}$ порождения источником букв a_j , $j = \overline{1, k}$, при заданных v предшест-

^{*)} Работа выполнена при поддержке Российского фонда фундаментальных исследований (грант № 09-07-00095-а) и Совета по грантам Президента РФ (ведущая научная школа № НШ-2121.2008.9)

вующих, $v \in A^s$, s любое целое неотрицательное число. Здесь, как и выше, при любом заданном v , $v \in A^s$ выполняется равенство:

$$\sum_{j=1}^k \theta_{vj} = 1; \quad s = 0, 1, 2, \dots$$

Множество всех стационарных источников обозначим Ω_∞ .

Если u – произвольное слово в алфавите A , то через $P_\theta(u)$ обозначим вероятность слова u , порождённого источником θ . Число $|u|$ букв в слове u назовём его длиной. Энтропию источника θ обозначим $H(\theta)$. Как известно [3, 4], если θ – стационарный источник, то

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

Здесь и в дальнейшем $\log x = \log_2 x$, $0 \log 0 = 0$.

Для бернуллиевского источника θ его энтропия $H_0(\theta)$ определяется равенством

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

Если θ – марковский источник с памятью s , то его энтропия $H_s(\theta)$ находится по формуле

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

где θ_{0v} – начальные стационарные вероятности слов v , $v \in A^s$. При $s=0$ из (3) получаем соотношение (1). Если θ – произвольный стационарный дискретный источник и $H(\theta)$ – его энтропия, то справедливо равенство [2, 3]

$$H(\theta) = \lim_{s \rightarrow \infty} H_s(\theta). \quad (4)$$

Рассмотрим T – конечное полное множество слов во входном алфавите. Множество T – полное, если оно префиксное и при любом непустом слове u (в алфавите A) множество слов $T \cup u$ уже не префиксное. Множество T назовём кодовым. Примером кодового множества может служить множество всех слов длины n взятых в алфавите A , т.е. A^n ; множество $A^n \setminus \underbrace{a_1, \dots, a_k}_n$, не является кодовым, потому что оно не полное.

Пусть θ – произвольный источник из Ω_s , T – произвольное кодовое множество. Обозначим через $\theta(T)$ марковскую цепь, состояниями которой являются слова из T , а переходные вероятности $P_{\theta(T)}(u/v)$, $u, v \in T$, индуцируются источником θ . Будем рассматривать только марковские источники с памятью s , переходные вероятности которых строго положительны. Тогда для любых $u, v \in T$ выполняются неравенства $P_{\theta(T)}(u/v) > 0$, поэтому для марковской цепи $\theta(T)$ существует стационарное распределение $P_{\theta(T)}^0(u) > 0$, $u \in T$. Средняя длина слова $d(T, \theta)$ для множества T , как доказано в [5], равна

$$d(T, \theta) = \sum_{u \in T} P_{\theta(T)}^0(u) |u|. \quad (5)$$

В этой же работе доказаны тождества Вальда, которые имеют вид

$$\sum_{u \in T} P_{\theta(T)}^0(u) \cdot r_v(u) = (d(T, \theta) - \hat{s} + 1) \theta_{0v}, \quad (6)$$

$$\sum_{u \in T} P_{\theta(T)}^0(u) \cdot r_{vi}(u) = (d(T, \theta) - s) \theta_{0v} \theta_{vi}, \quad (7)$$

где $r_v(u)$, $r_{vi}(u)$ – число вхождений блоков v, va_i , $v \in A^s$, в слово u , соответственно, $\hat{s} = \max(s, 1)$.

2. Полубесконечная последовательность букв, порожаемая 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 Мак-Милана – Крафта [3, 4] следует, что самое общее из всех возможных дешифрируемых кодирований φ такое, что 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 говорят, 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) кодирование, отображающее блоки в слова переменной длины (обозначается BV);
- 2) кодирование, отображающее слова переменной длины в блоки (обозначается VB);
- 3) кодирование, отображающее слова переменной длины в слова переменной длины (обозначается VV);
- 4) кодирование, отображающее блоки в слова переменной длины (обозначается 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 [5] доказано, что стоимость кодирования типа σ , $\sigma = BV, VB, VV$, для произвольного кодового множества T и любого источника θ , $\theta \in \Omega_s$, $0 \leq s < \infty$, определяется равенством

$$C_\sigma(T, \theta, \varphi) = \frac{1}{d_s(T, \theta) - \hat{s} + 1} \sum_{u \in T} P_{\theta(T)}^0(u) |\varphi(u)|. \quad (8)$$

Эффективность кодирования φ как обычно [1, 3, 4] будем оценивать разностью между стоимостью кодирования $C_\sigma(T, \theta, \varphi)$ и энтропией источника $H(\theta)$. Эта разность в дальнейшем называется избыточностью кодирования и обозначается $r_\sigma(T, \theta, \varphi)$, т.е.

$$r_\sigma(T, \theta, \varphi) = C_\sigma(T, \theta, \varphi) - H(\theta). \quad (9)$$

Избыточностью универсального кодирования типа σ для множества источников Ω и с заданной сложностью N назовём величину $R_\sigma(N, \Omega)$

$$R_\sigma(N, \Omega) = \inf_{\varphi} \sup_{\theta \in \Omega} r_\sigma(T, \theta, \varphi). \quad (10)$$

Здесь нижняя грань берётся по всем кодированиям φ , для которых кодовое множество T имеет не более чем k^N слов. Построение хорошего 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 Ω состоит из единственного источника, то мы имеем дело с кодированием известного источника, которое подробно изучено для различных типов кодирования, например, в работах [1, 3, 6–12]. Универсальное кодирование марковских источников различных типов также хорошо изучено. Подробную библиографию по этому вопросу можно найти в [13, 16]. Особо отметим работу В.Ф. Бабкина, Ю.М. Штарькова [14], в которой изучалось BV кодирование для стационарных источников. В частности,

в этой работе было доказано, что существует последовательность BV кодирований φ_N такая, что для любого стационарного источника θ избыточность кодирования $r_{BV}(A^N, \theta, \varphi_N)$ стремится к нулю. В тоже время легко показать, что это стремление к нулю не является равномерным по θ , более того при $N \rightarrow \infty$ избыточность универсального кодирования множества всех стационарных источников $R_{BV}(N, \Omega_\infty)$ стремится к бесконечности. Вопрос о равномерной сходимости $r_{BV}(A^N, \theta, \varphi_N)$ в [14] не исследовался. Кодирование, построенное в [14], получило название *слабоуниверсальное кодирование*. При построении слабоуниверсального BV кодирования основная сложность состоит в определении отображения φ_N , так как область определения при таком кодировании определена – это множество всех слов длины N в алфавите A . При построении кодирования типа VB основная трудность состоит в конструировании области определения кодирования φ_N , т.е. в определении кодового множества T_N .

В настоящей работе предложен метод универсального равномерного по выходу кодирования для множества марковских источников связности s , получена верхняя оценка избыточности, которая примерно в два раза лучше оценки автора [15]. Доказано существование слабоуниверсального кодирования типа BV для множества всех стационарных дискретных источников и сформулированы необходимые и достаточные условия существования универсального кодирования для произвольного множества источников Ω , $\Omega \in \Omega_\infty$.

2. Равномерное по выходу кодирование марковских источников

В этом параграфе будет предложен метод кодирования марковских источников с памятью s , получена оценка избыточности предложенного метода и доказана его универсальность. При доказательстве основного утверждения параграфа нам потребуются следующие понятия и обозначения. Марковский источник θ связности 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 слова u , $|u| > s$, начинающегося блоком v , $v \in A^s$, источником θ определяется равенством

$$P_\theta(u) = \theta_{0v} \prod_{v \in A^s} \prod_{i=1}^k \theta_{vi}^{r_{vi}(u)}. \quad (11)$$

На множестве источников Ω_s определим КТ – распределение $\omega(\theta)$ [13], которое задаётся формулой

$$\omega(\theta) = \left[\frac{r\left(\frac{k}{2}\right)}{k\pi^2} \right]^k \cdot \frac{1}{\sqrt{\prod_{v \in A^s} \prod_{i=1}^k \theta_{vi}}}, \quad (12)$$

Проинтегрировав вероятность слова u , порождённого источником θ , по множеству источников Ω_s , если на Ω_s задана плотность $\omega(\theta)$, получим [13]:

$$\overline{P}_s(u) = \left[\frac{\Gamma\left(\frac{k}{2}\right)}{k\pi^{\frac{k}{2}}} \right]^{ks} \prod_{v \in A^s} \frac{\prod_{j=1}^k \Gamma\left(r_{vj}(u) + \frac{1}{2}\right)}{\Gamma\left(r_v(u) + \frac{k}{2}\right)}. \quad (13)$$

где $\Gamma(z)$ – гамма-функция от z . Используя для функции $\Gamma(z)$ формулу Стирлинга в виде

$$\log \Gamma(z) = \log \sqrt{2\pi} + \left(z - \frac{1}{2}\right) \log \left(z - \frac{1}{2}\right) - z \log e + c(z) \log e, \quad (14)$$

где постоянная $c(z)$ удовлетворяет неравенствам $\frac{1}{2} \log(e-1) \leq c(z) \log e \leq \frac{1}{2} \log e$, из (13) получаем

$$-\log \overline{P}_s(u) = \sum_{v \in A^s} r_v(u) F_s(u) + \frac{k-1}{2} \sum_{v \in A^s} \log \hat{r}_v(u) + c. \quad (15)$$

Здесь $F_s(u)$ – квазиэнтропия u , определяемая равенством

$$F_s(u) = - \sum_{v \in A^s} \frac{r_v(u)}{|u| - s} \sum_{i=1}^k \frac{r_{vi}(u)}{r_v(u)} \log \frac{r_{vi}(u)}{r_v(u)}.$$

Сформулируем и докажем основное утверждение параграфа.

Теорема 1. Для любого фиксированного s , $0 \leq s < \infty$, существует последовательность кодирований φ_N типа VB , для которых избыточность кодирования $r_{VB}(T_N, \theta, \varphi_N)$ при любом источнике θ , $\theta \in \Omega_s$, удовлетворяет неравенству

$$r_{VB}(\varphi_N, \theta, T_N) \leq \frac{k^s(k-1)+2}{2} \cdot \frac{\log d(T_N, \theta) + c}{d_s(T_N, \theta)},$$

где постоянная c не зависит ни от θ , ни от T_N .

Доказательство. Как уже отмечалось ранее, каждое кодирование определяется тройкой $(T, \varphi, \varphi(T))$, где T – область определения T , $\varphi(T)$ – область значений отображения φ . Для равномерного по выходу кодирования область значения отображения φ определяется достаточно просто. В этом случае $\varphi(T) = B^{\lceil \log \|T\| \rceil}$, где $\lceil x \rceil$ – наименьшее целое, большее или равное x , $\|T\|$ – мощность множества T . Таким образом, при построении равномерных по выходу кодирований вся сложность заключается в построении кодовых множеств T_N , $N = 1, 2, \dots$

Зафиксируем произвольное натуральное N , в кодовое множество T_N включим все слова u , для которых выполняется неравенство

$$\frac{1}{\overline{P}_s(u)} \leq k^N \quad (16)$$

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

$$\frac{1}{\overline{P}_s(ua_j)} > k^N. \quad (17)$$

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

$$|\varphi_N(u)| = \lceil \log \|T_N\| \rceil. \quad (18)$$

Оценим избыточность предложенного метода кодирования. Из определения избыточности (9) и из (18) имеем

$$r_{VB}(T_N, \theta, \varphi_N) = \frac{\lceil \log \|T_N\| \rceil}{d_s(T_N, \theta) - \hat{s} + 1} - H(\theta). \quad (19)$$

Кодирование φ_N – дешифруемое, поэтому величина $r_{VB}(T_N, \varphi_N, \theta)$ неотрицательна. Найдём верхнюю оценку этой величины. Из соотношения (16) следует, что при любом u , $u \in T_N$, справедливо неравенство

$$\overline{P}_s(u) \geq \frac{1}{k^N}.$$

Просуммировав последнее неравенство по всем словам u из T_N и учитывая, что в силу полноты T_N выполняется равенство $\sum_{u \in T_N} \overline{P}_s(u) = 1$, получим:

$$k^N \geq \|T_N\|. \quad (20)$$

Из (19) с учётом (20) и (17) следует

$$\begin{aligned} r_{VB}(T_N, \theta, \varphi_N) &\leq \frac{\sum_{u \in T_N} P_{\theta(T_N)}^0(u) \log \|T_N\|}{d_s(T_N, \theta)} - H(\theta) + \frac{1}{d_s(T_N, \theta)} \leq \\ &\leq \frac{\sum_{u \in T_N} P_{\theta(T_N)}^0(u) \log k^N}{d_s(T_N, \theta)} - H(\theta) + \frac{1}{d_s(T_N, \theta)} \leq \\ &\leq \frac{-\sum_{u \in T_N} P_{\theta(T_N)}^0(u) \log \overline{P}_s(ua_j)}{d_s(T_N, \theta)} - H(\theta) + \frac{1}{d_s(T_N, \theta)}. \end{aligned} \quad (21)$$

Из определения средней вероятности $\overline{P}_s(ua_j)$ слова ua_j по множеству источников Ω_s , свойств гамма функции и (13) для слова u , заканчивающегося блоком v , справедливо равенство:

$$\overline{P}_s(ua_j) = \overline{P}_s(u) \frac{r_{vj}(u) + \frac{3}{2}}{r_v(ua_j) + \frac{k}{2}}.$$

Отсюда следует, что

$$-\log \bar{P}_s(ua_j) = -\log \bar{P}_s(u) + \log \left[\frac{r_v(ua_j) + \frac{k}{2}}{r_{vj}(u) + \frac{3}{2}} \right]. \quad (22)$$

Так как для любого слова u , $u \in T_N$, всегда выполняется неравенство

$$\frac{r_v(ua_j) + \frac{k}{2}}{r_{vj}(u) + \frac{3}{2}} \leq |u| - s + \frac{k}{2},$$

то из (22) вытекает соотношение

$$-\log \bar{P}_s(ua_j) \leq -\log \bar{P}_s(u) + \log \left(|u| - s + \frac{k}{2} \right).$$

Отсюда и из (21) получаем

$$r_{VB}(T_N, \theta, \varphi_N) \leq \frac{-\sum_{u \in T_N} P_{\theta(T_N)}^0(u) \log \bar{P}_s(u)}{d_s(T_N, \theta)} - H(\theta) + \frac{\sum_{u \in T_N} P_{\theta(T_N)}^0(u) \log \left(|u| - s + \frac{k}{2} \right) + 1}{d_s(T_N, \theta)}.$$

Значение $\log \bar{P}_s(u)$ определяется равенством (15), учитывая которое последнее неравенство можно переписать в виде

$$\begin{aligned} r_{VB}(T_N, \theta, \varphi_N) &\leq \frac{\sum_{v \in A^s} \sum_{u \in T_N} P_{\theta(T_N)}^0(u) r_v(u) F_s(u)}{d_s(T_N, \theta)} - H(\theta) + \\ &+ \frac{\frac{k-1}{2} \sum_{v \in A^s} \sum_{u \in T_N} P_{\theta(T_N)}^0(u) \log \hat{r}_v(u) + c}{d(T_N, \theta)} + \frac{\sum_{u \in T_N} P_{\theta(T_N)}^0(u) \log \left(|u| - s + \frac{k}{2} \right)}{d_s(T_N, \theta)}. \end{aligned} \quad (23)$$

Из определения квазиэнтропии $F_s(u)$ неравенства Иенсена для функции $-x \log x$, тождеств Вальда (6), (7) и определения величины $d(T_N, \theta)$, см. (5), получаем

$$\frac{\sum_{v \in A^s} \sum_{u \in T_N} P_{\theta(T_N)}^0(u) r_v(u) F_s(u)}{d_s(T_N, \theta)} - H(\theta) \leq 0. \quad (24)$$

Из неравенств Иенсена для функции $\log x$, тождеств Вальда (6), (7) и равенства (5) заключаем, что

$$\sum_{u \in T_N} P_{\theta(T_N)}^0 \log(|u| - s) \leq \log(d_s(T_N, \theta) - s). \quad (25)$$

Так как, по определению, $r_\alpha(u) \leq |u| - s$, то, используя (25), имеем

$$\sum_{u \in T_N} P_{\theta(T_N)}^0 \log \hat{r}_v(u) \leq \sum_{u \in T_N} P_{\theta(T_N)}^0 \log(|u| - s) \leq \log(d_s(T_N, \theta) - s). \quad (26)$$

Из (23) и соотношений (24), (25), (26) окончательно вытекает

$$r_{VB}(T_N, \theta, \varphi_N) \leq \frac{k^s(k-1)}{2} \cdot \frac{\log d(T_N, \theta)}{d(T_N, \theta)} + \frac{\log d(T_N, \theta) + c}{d(T_N, \theta)},$$

где c не зависит от θ . Теорема доказана.

Из доказанной теоремы следует, что для множества Ω_s марковских источников с памятью s , $0 \leq s < \infty$, существует универсальное равномерное по выходу кодирование. В самом деле, пусть

$$d_s(T_N) = \inf_{\theta \in \Omega_s} d_s(T_N, \theta).$$

Тогда справедливо следующее утверждение.

Следствие. Для избыточности $R_{VB}(N, \Omega_s)$ универсального равномерного по выходу кодирования с заданной сложностью N справедлива оценка

$$R_{VB}(N, \Omega_s) \leq \frac{k^s(k-1) + 2}{2} \cdot \frac{\log d_s(T_N)}{d_s(T_N)} + \frac{c}{d_s(T_N)}, \quad (27)$$

где c не зависит от θ , т.е. существует универсальное равномерное по выходу кодирование для множества источников Ω_s .

Доказательство. Утверждение следствия вытекает непосредственно из теоремы. Согласно определению величин $R_{VB}(N, \Omega_s)$ и $r(T_N, \theta, \varphi_N)$, имеем

$$R_{VB}(N, \Omega_s) \leq \sup_{\theta \in \Omega_s} r(T_N, \theta, \varphi_N). \quad (28)$$

По построению кодовое множество T_N содержит не более k^N кодовых слов. Учитывая, что при x , стремящемся к бесконечности, функция $\frac{\log x}{x}$ является убывающей, из неравенства (28) и теоремы 1 вытекает справедливость оценки (27). Правая часть (27) не зависит от θ и с ростом N стремится к нулю, потому что

$$d_s(T_N) \geq u(N) = \min_{u \in T_N} u,$$

а $u(N)$ стремится к бесконечности с ростом N . Таким образом, доказано, что избыточность $R_{VB}(N, \Omega_s)$ универсально равномерного по выходу кодирования стремится к нулю, т.е. для множества источников Ω_s существует универсальное кодирование. Следствие доказано.

3. Кодирование типа VB для стационарных источников

В этом параграфе будут доказаны основные утверждения работы. Имеет место утверждение.

Теорема 2. Для множества всех стационарных источников Ω_∞ существует слабо-универсальное равномерное по выходу кодирование.

Доказательство. Каждый стационарный источник θ , $\theta \in \Omega_\infty$, задается условными вероятностными распределениями $\theta_s(a_i | \nu)$, $a_i \in A$, $\nu \in A^s$, $s = 0, 1, 2, \dots$ появления буквы a_i после блока ν . Таким образом, каждый стационарный источник θ определяет последова-

тельность марковских источников $\theta_s, s = 0, 1, 2, \dots$, при s , стремящемся к бесконечности, энтропия $H(\theta_s)$ источника θ_s не возрастаая сходится к энтропии $H(\theta)$ источника θ , точнее, справедливы соотношения:

$$H(\theta_0) \geq H(\theta_1) \geq H(\theta_2) \geq \dots \geq H(\theta_s) \geq H(\theta_{s+1}) \geq \dots, \quad (29)$$

$$\lim_{s \rightarrow \infty} H(\theta_s) = H(\theta).$$

Для любого фиксированного $s, 0 \leq s < \infty$, определена стоимость кодирования $C_{VB}(T, \theta, \varphi)$ (см. (8)). Покажем, что стоимость кодирования $C(T_N^s, \theta, \varphi^s)$ для кодирования типа VB , предложенного ранее, при N и s , стремящимися к бесконечности, существует и равна энтропии источника $H(\theta)$. Для этого нам нужно установить, что избыточность кодирования $r_{VB}(T_N^s, \theta, \varphi_N^s)$ для стационарного источника $\theta, \theta \in \Omega_\infty$, стремится к нулю с ростом N и s .

Согласно определению (19) имеем

$$r_{VB}(T_N^s, \theta, \varphi_N^s) = \frac{\lceil \log \|T_N^s\| \rceil}{d_s(T_N^s, \theta_s) - \hat{s} + 1} - H(\theta) \quad (30)$$

или

$$r_{VB}(T_N^s, \theta, \varphi_N^s) = \left[\frac{\lceil \log \|T_N^s\| \rceil}{d_s(T_N^s, \theta_s) - \hat{s} + 1} - H(\theta_s) \right] + [H(\theta_s) - H(\theta)]. \quad (31)$$

В равенстве (31) первое слагаемое в правой части, согласно следствию из предыдущего параграфа, ограничено асимптотически сверху величиной

$$\frac{k^s(k-1)+2}{2} \cdot \frac{\log d_s(T_N^s)}{d_s(T_N^s)}. \quad (32)$$

Из соотношений (29) и (32) следует, что с ростом s второе слагаемое также стремится к нулю. Если выбрать

$$s = o\left(\log d_s(T_N^s) - \log \log u_s(T_N^s)\right).$$

то первое и второе слагаемые правой части равенства (31) стремятся к нулю одновременно с ростом N . Тогда из (30) вытекает, что

$$\lim_{N \rightarrow \infty} r_{VB}(T_N^s, \theta, \varphi_N^s) = 0,$$

т. е.

$$\lim_{N \rightarrow \infty} C(T_N^s, \theta_s, \varphi_N^s) = H(\theta).$$

Теорема доказана.

Из теоремы 2 следует, что существует кодирование, при котором для любого фиксированного источника θ из Ω_∞ его избыточность стремится к нулю. Однако это стремление не является равномерным по множеству источников Ω_∞ . Как доказано в теореме 1 и её

следствии, для множества марковских источников Ω_s с памятью s стремление к нулю избыточности является равномерным по Ω_s . Нижеследующее утверждение даёт ответ на вопрос о равномерной сходимости к нулю избыточности для произвольного множества источников $\Omega \subset \Omega_\infty$, т.е. о существовании универсального равномерного по выходу кодирования для подмножества множества источников Ω_∞ .

Теорема 3. *Для существования универсального равномерного по выходу кодирования для множества источников Ω необходимо и достаточно, чтобы при s , стремящемся к бесконечности, энтропия $H(\theta_s)$ сходилась равномерно по θ , $\theta \in \Omega$ к энтропии $H(\theta)$.*

Доказательство. **Необходимость.** Пусть $H(\theta_s)$ сходится равномерно по θ к $H(\theta)$ на множестве Ω , при $s \rightarrow \infty$. Согласно определению, для любой последовательности кодовых множеств $\{T_N^s\}$, $N = 1, 2, \dots$, $0 \leq s < \infty$, справедливо равенство

$$r(T_N^s, \theta, \varphi_N) = r(T_N^s, \theta_s, \varphi_N) + H(\theta_s) - H(\theta).$$

Так как $r(T_N^s, \theta_s, \varphi_N) \geq 0$, то из последнего равенства имеем

$$H(\theta_s) - H(\theta) \leq r(T_N^s, \theta, \varphi_N) = r(T_N^s, \theta_s, \varphi_N) + H(\theta_s) - H(\theta). \quad (33)$$

В качестве T_N^s возьмём кодовые множества, построенные при доказательстве теоремы 1. Согласно следствию, из (27) и (33) имеем:

$$r(T_N^s, \theta, \varphi_N) \leq \frac{k^s(k-1)+2}{2} \cdot \frac{\log d_s(T_N^s)}{d_s(T_N^s)} + \frac{c}{d(T_N^s)} + H(\theta_s) - H(\theta). \quad (34)$$

По условию теоремы, существует s_0 , такое, чтобы для всех $\theta \in \Omega$, при $s > s_0$ выполнялось неравенство

$$0 \leq H(\theta_s) - H(\theta) < \frac{\varepsilon}{2}. \quad (35)$$

Величина $\frac{(k^{s_0}(k-1)+2) \log d_{s_0}(T_N^{s_0}) + c}{2d_{s_0}(T_N^{s_0})}$ не зависит от θ и при $N \rightarrow \infty$ стремится к нулю,

следовательно, существует N_0 такое, что при $N > N_0$ эта величина меньше $\frac{\varepsilon}{2}$. Таким обра-

зом, при сделанных предположениях избыточность $r(T_N^s, \theta, \varphi_N)$ стремится к нулю равномерно относительно множества источников Ω .

Достаточность. Если $H(\theta_s) - H(\theta)$ не стремится к нулю равномерно по множеству Ω , то из (32), точнее, из нижней оценки (33), следует, что для любой последовательности кодовых множеств T_N^s избыточность $r(T_N^s, \theta, \varphi_N)$ не стремится к нулю равномерно по множеству Ω .

Теорема доказана.

Литература

1. Шеннон К. Математическая теория связи. Работы по теории информации и кибернетике. – М.: ИЛ., 1969. – С.243–332.
2. Хорошевский В.Г. Архитектура вычислительных систем. – М.: МГТУ им. Н.Э. Баумана, 2005.
3. Фано Р. Передача информации. Статистическая теория связи. – М.: Мир, 1965. – 440 С.

4. Галлагер Р. Теория информации и надёжная связь. – М.: Сов.радио, 1974. – 720 С.
5. Могульский А.А., Трофимов В.К. Тождество Вальда и стоимость кодирования для цепей Маркова // VII Всесоюзная конференция по теории кодирования и передачи информации. Доклады // Теория информации. – Москва-Вильнюс. 1978. – ч.І. – С.112–116.
6. Кричевский Р.Е. Связь между избыточностью кодирования и достоверностью сведений об источнике // Пробл. передачи информ. – 1968. – Т.4. №3. – С.48–57.
7. Гильберт Э.Н., Мур Э.Ф. Двоичные кодовые системы переменной длины // Кибернетический сборник. – М.: ИЛ, № 3, 1961. – С.103–141.
8. Ходак Г.Л. Оценки избыточности при пословном кодировании сообщений, порождаемых бернуллиевским источником // Пробл. передачи информ. – 1972. – Т.8. № 2. – С.21–32.
9. Khodak G.L. Coding of Markov Sources With Low Redundancy // Proc. of 2nd International Symp. On Inform. Theory Tsahkadzor, Armenia. USSR, 1971, Akademiai Kiado. Budapest. – 1973. – p.201–204.
10. Jelinek F., Shneider K. On Variable-Length to Block Coding // IEEE Trans. Inform. Theory. – 1972. – v.18, no.6. p. 756–774.
11. Трофимов В.К. Эффективное кодирование блоками слов различной длины, порождённых известным марковским источником // Обработка информации в системах связи. – Л.:ЛЭИС, 1985. – С.9–15.
12. Ziv J. Variable-to-Fixed Length Codes are Better than Fixed-to-Variable Length Codes for Markov Sources // IEEE Trans. Inform. Theory. – 1990. – v.36, no.4. – p. 861–863.
13. Krichevskii R.E., Trofimov V.K. The Performace of Universal Encoding // IEEE Trans. on Inform. Theory. – 1981. – v. IT-27, no. 2. – p.199–207.
14. Shtarkov Yu.M., Babkin V.F. Combinatorial Encoding for Discrete Stationary Sources // In 2nd Internat. Symp. On Inform. Theory Tsahkadzor, Armenia. USSR, 1971, Akademiai Kiado. Budapest. –1973. – p.249–256.
15. Трофимов В.К. Равномерное по выходу кодирование марковских источников при неизвестной статистике // Пятый международный симпозиум по теории информации. Доклады. Москва – Тбилиси, 1979. – ч.ІІ. – С.172–175.
16. Krichevsky R. Universal Compression and Retrieval. – Dordrecht/Boston/London: 1994. – 219 p.

Статья поступила в редакцию 15.05.2010;

Трофимов Виктор Куприянович

Профессор, доктор технических наук, декан факультета информатики и вычислительной техники, заведующий Кафедрой высшей математики Сибирского государственного университета телекоммуникаций и информатики. Область научных исследований – теория информации, кодирование информации, порождаемой дискретным источником, оценка сложности кодирования.

Тел.&факс: (383)269-82-70; e-mail: trofimov@sibsutis.ru

Weaklyuniversal variable-to-fixed length codes of stationary sources.

V.K. Trofimov

Weaklyuniversal VB coding is proved to be for stationary sources set. Necessary and sufficient conditions of existing universal VB coding of stationary sources set owe obtained.

Keywords: entropy, coding, redundancy, capacity, source.