ПОСЛОВНОЕ КОДИРОВАНИЕ СООБЩЕНИЙ, ПОРОЖДЕННЫХ СТАЦИОНАРНЫМ ИСТОЧНИКОМ С НЕИЗВЕСТНОЙ СТАТИСТИКОЙ

В.К.Трофимов, В.И.Агульник, И.И.Резван

Доказано, что для любой последовательности кодовых множеств, у которых длина минимального слова стремится к нулю, существует слабоуниверсальное пословное кодирование для множества всех стационарных дискретных сообщений.

Ключевые слова: энтропия, кодирование, избыточность, источник сообщений

Основные определения. Постановка задачи

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

Пусть буквы конечного алфавита $A=\{a_1,a_2,...,a_k\}$, $2\le k<\infty$, порождаются источником θ. Мера, заданная на последовательности букв, порождаемой источником, определяет тип источника. Если вероятности $P_\theta\left(a_j\right), j=\overline{1,k}$ порождения букв $a_j, j=\overline{1,k}$ независимы и равны $\theta_j, j=\overline{1,k}$, $\theta_1+\theta_2+\ldots+\theta_k=1$, то источник называют бернуллиевским. Если же вероятность $P_\theta\left(a_i/a_j\right)$ появления очередной буквы a_i зависит от предыдущей буквы a_j, то положим $P_\theta\left(a_i/a_j\right)=\theta_{ji}$, $\sum_{i=1}^{k}\theta_{ji}=1$, $j=\overline{1,k}$, и в этом случае источник называют марковским. Если вероятность появления очередной буквы зависит от s предшествующих букв, т.е. $P_\theta\left(a_j/\nu\right)=\theta_{\nu j}$, где $\nu\in A^s$, то источник θ называют марковским с памятью s. Следует отметить, что для любого слова $\nu\in A^s$, $0\le s<\infty$, выполняется равенство $\sum_{j=1}^{k}\theta_{\nu j}=1$. Множество всех марковских источников с памятью s обозначим Ω_s. Дискретный стационарный источник *θ* задаётся всеми условными распределениями вероятностей $P_\theta\left(a_j/\nu\right)=\theta_{\nu j}$ порождения источником букв a_j, $j=\overline{1,k}$, при заданных ν предшествующих букв, $\nu\in A^s$, s - любое целое неотрицательное число. Здесь, как и выше, при любом заданном ν, $\nu\in A^s$ выполняется равенство:

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

Множество всех стационарных источников обозначим Ω_∞.

Пусть u – произвольное слово в алфавите A. Обозначим через $P_\theta(u)$ вероятность слова u, порожденного источником θ. Энтропию источника *θ* обозначим $H(\theta)$. Как известно [2-4], если *θ* – стационарный источник, то

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

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

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

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

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

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

где $\theta_{0\nu}$ – начальные стационарные вероятности слов ν, $\nu\in A^S$. При $s=0$ из (3) получаем соотношение (2). Если θ – произвольный стационарный дискретный источник и $H(\theta)$ его энтропия, то справедливо равенство [2–4]

$$H(\theta)=\lim_{s\to\infty}H_s(\theta)\,. \tag{4}$$

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. Такое множество T назовем кодовым. Примером кодового множества может служить множество всех слов длины n взятых в алфавите A, т.е. A^n; множество $A^n \setminus \underbrace{a_1, \ldots, a_k}_{n}$, 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 $P_{\theta(T)}(u/\nu)$, $u, \nu \in T$, 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 $P_{\theta(T)}(u/\nu) > 0$, 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 [5], равна

$$d_s(T, \theta) = \sum_{u \in T} P^0_{\theta(T)}(u) \cdot |u| , \qquad (5)$$

где $|u|$ - число букв в слове и (длина слова и).

В этой же работе доказаны тождества Вальда, которые имеют вид

$$\sum_{u \in T} P^0_{\theta(T)}(u) \cdot r_\nu(u) = (d_s(T, \theta) - \hat{s} + 1)\theta_{0\nu}, \quad (6)$$

$$\sum_{u \in T} P^0_{\theta(T)}(u) \cdot r_{\nu i}(u) = (d_s(T, \theta) - s)\theta_{0\nu}\theta_{\nu i}, \quad (7)$$

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

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 Мак-Милана-Крафта [2–4] 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 (VB);
3) кодирование, отображающее слова переменной длины в слова переменной длины (VV);
4) кодирование, отображающее блоки в слова переменной длины (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, назовем 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, $\theta \in \Omega_s$, $0 \le s < \infty$, определяется равенством:

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

Эффективность кодирования φ, как обычно [1–4], 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.

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

Избыточностью универсального кодирования типа σ для множества источников Ω с заданной сложностью N, назовем величину $R_\sigma(N, \theta)$:

$$R_\sigma(N, \theta) = \inf_{\varphi} r_\sigma(T, \theta, \varphi). \qquad (10)$$

Здесь нижняя грань берется 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 информации, порожденной известным источником, которое подробно изучено для различных типов кодирования, например, в работах [1–4, 6–13]. Универсальное кодирование марковских источников различных типов также хорошо изучено. Подробную библиографию по этому вопросу можно найти в [14 – 18]. Особо отметим работу В.Ф.Бабкина, Ю.М. Штарькова [15], 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 $N \to \infty$ 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 [15] не исследовался. Кодирование, построенное в [15], 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 [16] предложен метод универсального равномерного по выходу кодирования для множества марковских источников связанности s, получена верхняя оценка избыточности, которая примерно в два раза лучше оценки автора [17]. Доказано существование слабоуниверсального кодирования типа BV для множества всех стационарных источников. Сформулированы необходимые и достаточные условия, которым должно удовлетворять множество источников Ω, для того, чтобы на множестве Ω существовало универсальное кодирование.

В настоящей работе установлено, что для любой последовательности кодовых множеств $\{T_N\}, N = 1, 2, \ldots$, у которых $\min_{u \in T_N} u$ стремится к бесконечности, существует слабоуниверсальное пословное кодирование для множества стационарных источников. Получены необходимые и достаточные условия существования универсального кодирования для подмножества множества стационарных источников.

Неравномерное по выходу по входу и выходу универсальное кодирование марковских источников

В этом параграфе будет предложен метод VV 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 связанности s задается 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

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

На множестве источников Ω_s определим КТ распределение $\omega(\theta)$ [14], которое задается формулой

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

Проинтегрировав вероятность слова u, порожденного источником θ, по множеству источников Ω_s, если на Ω_s задана плотность $\omega(\theta)$, получим [14]:

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

Здесь $\Gamma(z)$ - гамма функции 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 $\frac{1}{2}\log(e-1) \le c(z)\log e \le \frac{1}{2}\log e$, из (13) получаем:

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

Здесь $F_S(u)$ - квазиэнтропия u, определяемая равенством

$$F_S(u) = -\sum_{v\in A^S}\frac{r_V(u)}{|u|-s}\sum_{i=1}^{k}\frac{r_{Vi}(u)}{r_V(u)}\log\frac{r_{Vi}(u)}{r_V(u)}. \quad (17)$$

Сформулируем и докажем основное утверждение параграфа.

Теорема 1. Для любого фиксированного s, $0 \le s < \infty$, для любой последовательности кодовых множеств в $\{T_N^\theta\}$, $N = 1, 2, \cdots$ такой, что $\min_{u\in T_N}|u|$ стремится к бесконечности с ростом N, существует последовательность универсальных кодирований $\{\varphi_N\}, N = 1, 2, \ldots$, для которых избыточность кодирования $r_{VB}(T_N, \theta, \varphi_N)$ при любом θ, $\theta \in \Omega_s$ удовлетворяет неравенству

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

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

Доказательство. Пусть $\theta \in \Omega_s$ произвольный марковский источник с памятью s, $0 \le s < \infty$, и $\{T_N\}$ - последовательность кодовых множеств, удовлетворяющая условию теоремы. Рассмотрим кодирование φ_N, которое каждому слову $u \in T_N$ ставит в соответствие слово $\varphi_N(u)$ длины

$$|\varphi_N(u)| = \left]-\log \overline{P}_s(u)\right[, \; u \in T_N$$

где $\overline{P}_s(u)$ определяется равенством (13).

Для слов с длинами кодовых слов $|\varphi_N(u)|, u \in T_N$, определенных равенствами (16), выполняется неравенство Крафта [2-4] и, следовательно, дешифруемое кодирование с такими длинами кодовых слов существует. Оценим избыточность предложенного кодирования. Из определения избыточности (9) и из (18) имеем:

$$r_{VV}(T_N, \theta, \varphi_N) = \frac{\sum_{u\in T_N} P^0_{\theta(T)}(u)\cdot \left]\log \overline{P}_s(\theta)\right[- H(\theta)}{d_s(T_N,\theta)}. \quad (19)$$

Учитывая, что $]x[< x+1$, из (19) получаем:

$$r_{VV}(T_N, \theta, \varphi_N) = \frac{-\sum_{u\in T_N} P^0_{\theta(T_N)}(u)\cdot\log \overline{P}_s(u)}{d_s(T_N,\theta)} - \\ -H(\theta) + \frac{1}{d_s(T_N,\theta)}. \quad (20)$$

Учитывая (15), последнее неравенство можно переписать в виде:

$$r_{VV}(T_N, \theta, \varphi_N) \le \frac{\sum_{v\in A^s}\sum_{u\in T_N} P^0_{\theta(T_N)}(u)\cdot r_v(u)\cdot F_s(u)}{d(T_N,\theta)} - \\ -H(\theta) + + \frac{\frac{k-1}{2}\sum_{v\in A^s}\sum_{u\in T_N} P^0_{\theta(T_N)}(u)\cdot\log\hat{r}_\alpha(u) + c}{d(T_N,\theta)} + \\ + \frac{1}{d_s(T_N,\theta)}. \quad (21)$$

Из определения квазиэнтропии $F_s(u)$ (17) неравенства Иенсена для функции $-x\log x$ и тождеств Вальда (6), (7) и определения величины $d(T_N, \theta)$, см. (5), получаем:

$$\frac{\sum_{v\in A^s}\sum_{u\in T_N} P^0_{\theta(T_N)}(u)\cdot r_v(u)\cdot F_s(u)}{d(T_N,\theta)} - H(\theta) \le 0\cdot \quad (22)$$

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

$$\sum_{u\in T_N} P^0_{\theta(T_N)}\log(|u| - s) \le \log\left(d_s(T_N,\theta) - s\right). \quad (23)$$

Так как, по определению $r_\alpha(u) \le |u| - s$, то с учетом (23) имеем:

$$\sum_{u\in T_N} P^0_{\theta(T_N)} \log \hat r_\nu(u) \le \sum_{u\in T_N} P^0_{\theta(T_N)} \log\left(|u|-s\right) \le \le \log\left(d_s(T_N,\theta)-s\right). \quad (24)$$

Из (21) с учетом соотношений (22), (23), (24) окончательно вытекает

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

где c не зависит от θ. Теорема доказана.

Из доказанной теоремы следует, что для множества Ω_S марковских источников связанности s, $0\le s<\infty$, существует универсальное неравномерное по выходу и по входу кодирование. В самом деле, пусть

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

Тогда справедливо следующее утверждение.

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

$$R_{VV}(N,\Omega_s) \le \frac{k^s(k-1)}{2}\cdot\frac{\log d_s(T_N)}{d_s(T_N)} + \frac{c}{d_s(T_N)}, \quad (25)$$

здесь 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:

$$R_{VV}(N,\Omega_s) \le \sup_{\theta\in\Omega_s} r_{VV}(T_N,\varphi_N,\theta). \quad (26)$$

Учитывая, что при x, стремящемся к бесконечности, функция $\frac{\log x}{x}$ является убывающей, из неравенства (26) и теоремы 1 вытекает справедливость оценки (25). Правая часть (25) не зависит от θ и с ростом N стремится к нулю, потому что

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

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

Кодирование типа VV для стационарных источников

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

Теорема 2. Для множества всех стационарных источников Ω_∞ существует слабоуниверсальное равномерное по выходу и входу кодирование.

Доказательство. Каждый стационарный источник θ, $\theta\in\Omega_\infty$, задается условными вероятностными распределениями $\theta_s(a_i|\nu)$, $a_i\in A$, $\nu\in A^s$, где $s=0,1,2,\ldots$ появления буквы a_i после блока ν. 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:

$$H(\theta_0)\ge H(\theta_1)\ge\ldots\ge H(\theta_s)\ge H(\theta_{s+1})\ge\ldots \quad (27)$$

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

Согласно определению (18) имеем

$$r_{VV}(T_N^s,\theta,\varphi_N^s) = c_{VV}(T_N^s,\theta,\varphi_N^s) - H(\theta) \quad (28)$$

или

$$r_{VV}(T_N^s,\theta,\varphi_N^s) = \left(c_{VV}(T_N^s,\theta,\varphi_N^s) - H(\theta)\right) + H(\theta_s) - H(\theta). \quad (29)$$

В равенстве (29) первое слагаемое в правой части, согласно следствию из предыдущего параграфа, ограничено асимптотически сверху величиной

$$\frac{k^s(k-1)}{2}\cdot\frac{\log d_s(T_N^s)}{d_s(T_N^s)}. \quad (30)$$

Из соотношения (27) и (30) следует, что с ростом s второе слагаемое также стремится к нулю. Если выбрать

$$s = o\left(\log d_s\left(T^s_N\right) - \log\log u_s\left(T^s_N\right)\right),$$

то первое и второе слагаемые правой части равенства (29) стремятся к нулю с ростом N. Тогда из (28) вытекает, что

$$\lim_{N\to\infty} r_{VV}\left(T^s_N, \theta, \varphi^s_N\right) = 0,$$

т.е.

$$\lim_{N\to\infty} C_{VV}\left(T^s_N, \theta, \varphi^s_N\right) = H(\theta).$$

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

Теорема 3. Для существования универсального неравномерного по выходу и входу 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 $s \to \infty$. Согласно определению, для любой последовательности кодовых множеств $\left\{T^s_N\right\}, N = 1, 2, \ldots,$ $0 \le s < \infty$ справедливо равенство

$$r_{VV}\left(T^s_N, \theta, \varphi_N\right) = r_{VV}\left(T^s_N, \theta_s, \varphi_N\right) + \left(H(\theta_s) - H(\theta)\right).$$

Так как $r_{VV}\left(T^s_N, \theta_s, \varphi_N\right) \ge 0$, то из последнего равенства имеем

$$\left(H(\theta_s) - H(\theta)\right) \le r_{VV}\left(T^s_N, \theta, \varphi_N\right) = \\ = r_{VV}\left(T^s_N, \theta_s, \varphi_N\right) + \left(H(\theta_s) - H(\theta)\right). \quad (31)$$

Согласно следствию, из (25) и (31) имеем:

$$r_{VV}\left(T^s_N, \theta, \varphi_N\right) \le \frac{k^s(k-1)}{2} \cdot \frac{\log d_s(T^s_N)}{d_s(T^s_N)} + \\ + \frac{c}{d(T^s_N)} + \left[H(\theta_s) - H(\theta)\right]. \quad (32)$$

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

$$0 \le H(\theta_s) - H(\theta) < \frac{\varepsilon}{2}. \quad (33)$$

Величина $\dfrac{k^{s_0}(k-1)\log d_{s_0}\left(T^{s_0}_N\right) + c}{2d_{s_0}\left(T^{s_0}_N\right)}$ не зависит от θ, и при $N \to \infty$ стремится к нулю, следовательно, существует N_0 такое, что при $N > N_0$ эта величина меньше $\frac{\varepsilon}{2}$. Таким образом, при сделанных предположениях избыточность $r_{VV}\left(T^s_N, \theta, \varphi_N\right)$ 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 (29), точнее, из нижней оценки (31), следует, что для любой последовательности кодовых множеств T^s_N избыточность $r_{VV}\left(T^s_N, \theta, \varphi_N\right)$ не стремится к нулю равномерно по множеству Ω. Теорема доказана.

СПИСОК ЛИТЕРАТУРЫ

1. Шеннон, К. Математическая теория связи. Работы по теории информации и кибернетике/ К.Шеннон – М.: 1969. С.243–332.
2. Яглом А.М. Вероятность и информация/ А.М.Яглом, Яглом И.М. – М.: Наука, 1973,- 511 с.
3. Фано, Р. Передача информации. Статистическая теория связи./ Р.Фано – М.: 1965, 440 с.
4. Галлагер, Р. Теория информации и надёжная связь./ Р.Галагер – М.: 1974. – 720 с.
5. Могульский, А.А. Тождество Вальда и стоимость кодирования для цепей Маркова. // А.А.Могульский, В.К.Трофимов. VII Всесоюзная конференция по теории кодирования и передачи информации. Доклады // Теория информации. – Москва-Вильнюс. 1978., ч.I. С.112–116.

6. Кричевский, Р.Е. Связь между избыточностью кодирования и достоверностью сведений об источнике // Р.Е.Кричевский. Пробл. передачи информ. 1968. Т.4. №3. С.48–57.
7. Гильберт, Э.Н. Двоичные кодовые системы переменной длины. // Э.Н.Гильберт, Э.Ф.Мур. Кибернетический сборник. – М.: 1961, № 3, С.103–141.
8. Ходак, Г.Л. Оценки избыточности при пословном кодировании сообщений, порождаемых бернуллиевским источником. // Г.Л.Ходак. Пробл. передачи информ. – 1972. Т.8. № 2. С.21–32.
9. Khodak, G.L. Coding of Markov Sources With Low Redundancy // G.L.Khodak. Proc. of 2nd International Symp. On Inform. Theory Tsahkadzor, Armenia. USSR, 1971, Akademiai Kiado. Budapest. 1973. P.201–204.
10. Jelinek, F. On Variable-Length to Block Coding // F.Jelinek, K.Shneider. IEEE Trans. Inform. Theory. –1972. V.18, №.6. P.756–774.
11. Трофимов, В.К. Эффективное кодирование блоками слов различной длины, порождённых известным марковским источником // В.К.Трофимов. Обработка информации в системах связи. – Л.: 1985. С.9–15.
12. Ziv, J. Variable-to-Fixed Length Codes are Better than Fixed-to-Variable Length Codes for Marcov Sources // J.Ziv. IEEE Trans. Inform. Theory. 1990. V.36. №.4. P.861–863.
13. Трофимов В.К. Неравномерное по входу кодирование сообщений, порожденных стационарным источником// Трофимов В.К., Агульник В.И., Резван И.И. Ползуновский вестник. Измерение, информатизация, моделирование: проблемы и перспективы технологий разработки и применения, №3/1, Барнаул, 2011. С.224-229.
14. Krichevskii, R.E. The Performace of Universal Encoding // R.E.Krichevskii, V.K.Trofimov. IEEE Trans. on Inform. Theory. 1981. V.IT-27. №2. P.199–207.
15. Shtarkov, Yu.M. Combinatorial Encoding for Discrete Stationary Sources // Yu.M.Shtarkov,. V.F.Babkin. Proc. of 2nd International Symp. On Inform. Theory Tsahkadzor, Armenia. USSR, 1971, Akademiai Kiado. Budapest. 1973., P.249–256.
16. Трофимов В.К. Слабоуниверсальное равномерное по выходу кодирование // В.К.Трофимов. Вестник СибГУТИ, 2010. №2, С.101-111.
17. Трофимов В.К. Равномерное по выходу кодирование марковских источников при неизвестной статистике// В.К.Трофимов. V международный симпозиум по теории информации. Доклады. Москва – Тбилиси, 1979. ч.II, С.172–175.
18. Krichevsky, R. Universal Compression and Retrieval.// R.Krichevskii. Dordrecht/Boston/London: 1994. P.219.

Д.т.н., профессор, декан факультета информатики и вычислительной техники ***Трофимов В.К.****, тел. (383) 269-82-70, e-mail: trofimov@sibsutis.ru; и.о.доцента Агульник В.И., тел. (383) 269-82-71, e-mail: agulnik@sibsutis.ru; к.т.н., доцент Резван И.И., e-mail: rezvan@rambler.ru; Сибирский государственный университет телекоммуникаций и информатики (г. Новосибирск).*

УДК 681.518

АЛГОРИТМЫ ОБРАБОТКИ ДАННЫХ В ГЕОИНФОРМАЦИОННОЙ СИСТЕМЕ ДЛЯ УЧЕТА ЗЕМЕЛЬНЫХ УЧАСТКОВ

С.В. Еремеев

В статье рассмотрены вопросы, связанные с разработкой и реализацией алгоритмов построения и учета земельных участков. Показана возможность загрузки координат из внешнего файла, на основе которых осуществляется построение земельного участка на карте. На основании сформулированных правил в виде топологических отношений между слоями производится учет объектов, расположенных на земельном участке. Разработан программный модуль для внедрения в муниципальную ГИС

Ключевые слова: геоинформационная система, земельные участки, кадастровый паспорт, топологические отношения

Введение

В последнее время в связи с динамическим развитием городской инфраструктуры у муниципальных служб и граждан возрастает потребность в получении актуальной картографической информации с ее привязкой к местности. Все более очевидной и актуальной становится задача создания и широкого использования единой общегородской справочно-картографической информационной системы, которая могла бы удовлетворить большинство запросов разнообразных пользователей. Для решения задач эффективного хранения, оперативного извлечения данных об объектах городской территории и выполнения с ними аналитических операций наи-