Math-Net.Ru
All Russian mathematical portal

Yu. M. Shtar'kov, Universal Sequential Coding of Single Messages, *Probl. Peredachi Inf.*, 1987, Volume 23, Issue 3, 3–17

Use of the all-Russian mathematical portal Math-Net.Ru implies that you have read and agreed to these terms of use
http://www.mathnet.ru/eng/agreement

Download details:
IP: 77.188.22.180
May 1, 2017, 12:43:19

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том XXIII 1987 Вып. 3

УДК 621.391.15

УНИВЕРСАЛЬНОЕ ПОСЛЕДОВАТЕЛЬНОЕ КОДИРОВАНИЕ ОТДЕЛЬНЫХ СООБЩЕНИЙ

Штарьков Ю. М.

Введено понятие избыточности кодирования отдельных сообщений. Предложен метод кодирования, обеспечивающий равномерное ограничение этой избыточности для всех сообщений на выходе источника с неизвестной статистикой. Для этого же критерия исследованы возможности последовательного кодирования сообщений, в том числе и на выходе источников с неизвестной статистикой. Получены верхние оценки избыточности для источников без памяти, марковских цепей и различных множеств марковских источников.

§ 1. Введение

Пусть A и B — дискретные алфавиты из $m\geqslant 2$ букв и $d\geqslant 2$ символов соответственно; A^k, $k=1, 2,\ldots$ — множество всех m^k последовательностей $\alpha^k=\alpha_1\ldots\alpha_k$ из k букв алфавита A; $p(\alpha^k|\omega)$ — вероятность появления α^k на выходе источника ω; φ_n — однозначно декодируемый d-ичный код, каждому блоку $\alpha^n\in A^n$ ставящий в соответствие слово $\varphi(\alpha^n)$ из $L(\alpha^n|\varphi_n)$ символов алфавита B. «Средняя» избыточность кодирования источника ω блоковым кодом φ_n определяется величиной

$$r_n(\varphi_n, \omega)=n^{-1}\{E_\omega[L(\alpha^n|\varphi_n)]-H_\omega(A^n)\} \qquad (1)$$

или

$$r_n^*(\varphi_n, \omega)=r_n(\varphi_n, \omega)+[n^{-1}H_\omega(A^n)-H_\omega]\geqslant r_n(\varphi_n, \omega),$$

где $E_\omega[L(\alpha^n|\varphi_n)]$ — среднее значение $L(\alpha^n|\varphi_n)$ на распределении $\{p(\alpha^n|\omega), \alpha^n\in A^n\}$, $H_\omega(A^n)$ — d-ичная энтропия этого распределения, а $H_\omega=\lim\,[n^{-1}\times \times H_\omega(A^n)]$, $n\to\infty$. H_ω и соответственно $r_n^*(\varphi_n, \omega)$ определены для всех стационарных и некоторых нестационарных источников, а $r_n(\varphi_n, \omega)$ — для произвольных ансамблей сообщений.

В задачах универсального кодирования источников с неизвестной статистикой обычно рассматривается множество Ω источников, для каждого элемента ω которого известны распределения $\{p(\alpha^k|\omega), \alpha^k\in A^k, k=1, 2,\ldots\}$, а эффективность кода характеризуется величиной $r_n^*(\varphi_n, \Omega)=\sup\{r_n^*(\varphi_n, \omega), \omega\in\Omega\}$ и ее отличием от $r_n^*(\Omega)=\inf\{r_n^*(\varphi_n, \Omega), \varphi_n\in\Phi_n\}$, где Φ_n — множество всех однозначно декодируемых кодов с m^n или более словами. Аналогично вводятся $r_n(\varphi_n, \Omega)$ и $r_n(\Omega)$. Последовательность кодов $\{\varphi_n, n=1, 2,\ldots\}$ называется универсальной для Ω, если $r_n^*(\varphi_n, \Omega)$ стремится к нулю с ростом n, и асимптотически оптимальной, если при этом отношение $r_n^*(\varphi_n, \Omega)/r_n^*(\Omega)$ стремится к 1.

В настоящей работе используется критерий максимальной (а не средней) избыточности [1—3]. Применительно к этому критерию исследуются возможности последовательного кодирования (см., например, [4—6]) в условиях как известной, так и неизвестной статистики.

Ниже $\log x=\log_d x$, $\exp x=d^x$, $\lceil x\rceil$ — минимальное целое число, не меньшее x, $\alpha^i\beta^j$ — последовательная (без пропуска) запись $\alpha^i\in A^i$ и $\beta^j\in A^j$.

§ 2. Избыточность кодирования отдельных сообщений [1]

Величину $-\log p(\alpha^n|\omega)$ естественно трактовать как количество информации, содержащейся в блоке α^n на выходе источника ω (см., например, [7]). Тогда

$$(2)\qquad \rho(\alpha^n|\varphi_n, \omega)=n^{-1}[L(\alpha^n|\varphi_n)+\log p(\alpha^n|\omega)]$$

— *избыточность кодирования блока* α^n *на выходе источника* ω *кодом* φ_n. В случае, когда ω — произвольный элемент множества Ω, введем

$$(3)\qquad \rho(\alpha^n|\varphi_n, \Omega)=\sup\{\rho(\alpha^n|\varphi_n, \omega), \omega\in\Omega\}= =n^{-1}\{L(\alpha^n|\varphi_n)+\log p(\alpha^n|\Omega)\},$$

где

$$(4)\qquad p(\alpha^n|\Omega)=\sup\{p(\alpha^n|\omega), \omega\in\Omega\}.$$

Будем оценивать кодирование максимумом «индивидуальной» избыточности

$$(5)\qquad \rho_n(\varphi_n, \Omega)=\max\{\rho(\alpha^n|\varphi_n, \Omega), \alpha^n\in A^n\}$$

и его отличием от

$$(6)\qquad \rho_n(\Omega)=\inf\{\rho_n(\varphi_n, \Omega), \varphi_n\in\Phi_n\}$$

(для источников с известной статистикой $\Omega=\omega$, $\rho_n(\varphi_n, \Omega)$ — максимум (2), а $\rho_n(\Omega)$ меньше $1/n$). Согласно определению $r_n(\varphi_n, \Omega)$, (1), (2) и (5), для любых кода φ_n и множества Ω

$$r_n(\varphi_n, \Omega)=\sup\{E_\omega[\rho(\alpha^n|\varphi_n, \omega)], \omega\in\Omega\}\leqslant\rho_n(\varphi_n, \Omega)$$

и соответственно

$$(7)\qquad r_n(\Omega)\leqslant\rho_n(\Omega).$$

Теорема 1. *Для любого множества источников* Ω

$$(8)\qquad \rho_n(\Omega)\geqslant n^{-1}\log S_n(\Omega),$$

где

$$(9)\qquad S_n(\Omega)=\sum_{\alpha^n\in A^n} p(\alpha^n|\Omega),$$

и существует однозначно декодируемый код $\varphi_n^*(\Omega)$ *с длинами кодовых слов*

$$(10)\qquad L[\alpha^n|\varphi_n^*(\Omega)]=\lceil\log S_n(\Omega)-\log p(\alpha^n|\Omega)\rceil,$$

для которого

$$(11)\qquad \rho_n(\varphi_n^*, \Omega)<n^{-1}[\log S_n(\Omega)+1]\leqslant\rho_n(\Omega)+1/n.$$

Доказательство. Согласно (4) и (8),

$$(12)\qquad q_n(\alpha^n|\Omega)=p(\alpha^n|\Omega)/S_n(\Omega)$$

— распределение вероятностей на A^n. Поскольку множество длин кодовых слов любого однозначно декодируемого кода φ_n удовлетворяет неравенству Крафта, найдется по крайней мере один блок α^n, для которого $\exp\{-L(\alpha^n|\varphi_n)\}\leqslant q_n(\alpha^n|\Omega)$, и из этого неравенства, (3), (5) и (12) сразу следует (8). С другой стороны, множество длин кодовых слов (10) удовлетворяет неравенству Крафта и из (10), (3) и (5) сразу получаем (11). Теорема доказана.

Код $\varphi_n^*(\Omega)$, удовлетворяющий (10), назовем кодом максимальных вероятностей или МВ-кодом (в [1—3] использовался термин «коды макси-

[1] Основное содержание этого параграфа изложено в работах [1—3].

мального правдоподобия»). Он определен для любых n и произвольных множеств источников — стационарных и нестационарных, параметрических и непараметрических, задаваемых списком, и т. д. Если Ω содержит единственный элемент, то $S_n(\Omega)=1$ и МВ-код совпадает с хорошо известным кодом Шеннона — Фано.

Пусть Ω — параметрическое множество, задаваемое областью Θ значений неизвестного векторного параметра θ, а $\hat\theta(\alpha^n)$ — оценка максимального правдоподобия для θ (с учетом области Θ). Тогда $p(\alpha^n|\omega)=p(\alpha^n|\theta)$ и $p(\alpha^n|\Omega)=p[\alpha^n|\hat\theta(\alpha^n)]$. Сумма $p(\alpha^n|\hat\theta(\alpha^n))$ по всем α^n с заданным $\hat\theta(\alpha^n)=\hat\theta_0$ равна $p_n(\hat\theta_0|\Theta)$ — максимальной вероятности получения оценки $\hat\theta_0$. Поэтому

$$S_n(\Omega)=S_n(\Theta)=\sum p_n(\hat\theta_0|\Theta)\leqslant\mathcal{N}(n,\Theta) \tag{13}$$

(суммирование выполняется по всем $\mathcal{N}(n,\Theta)$ различным оценкам, существующим при заданных n и Θ) и

$$q_n(\alpha^n|\Omega)=\frac{p_n(\hat\theta_0|\Theta)}{S_n(\Theta)}\frac{p(\alpha^n|\hat\theta_0)}{p_n(\hat\theta_0|\Theta)},$$

где $\hat\theta_0=\hat\theta(\alpha^n)$, первый сомножитель — вероятность вида (12) для оценки $\hat\theta_0$ (см. (13)), а второй — вероятность появления α^n при условии, что задана оценка $\hat\theta_0$. Таким образом, кодовое слово МВ-кода можно представить в виде последовательности кодового слова МВ-кода для оценки $\hat\theta(\alpha^n)$ и кодового слова для описания α^n при заданном $\hat\theta$. Последнее особенно удобно, если для θ существует достаточная статистика и соответственно второй сомножитель не зависит от Θ. Заметим, что достаточные статистики неоднократно использовались при построении универсальных (по критерию $r_n^*(\varphi_n,\Omega)$) кодов.

Приведем примеры вычисления и получения оценок $S_n(\Omega)$ сверху. Отметим, что оценка этой суммы с точностью до постоянного множителя $c>1$ увеличивает оценку $\rho_n(\Omega)$ менее чем на $(1+\log c)/n$.

Пример 1. Множеству всех m-ичных источников без памяти соответствует $(m-1)$-мерный симплекс Θ_0^m распределений $\theta=(\theta_1,\ldots,\theta_m)$ на A, для которого

$$S_n(\Theta_0^m)=\sum_{k=1}^{m}C_m^kS_n^{(k)},\qquad n\geqslant m, \tag{14}$$

где

$$S_n^{(k)}=\sum_{i=1}^{n-k+1}C_n^i\left(\frac{i}{n}\right)^i\left(1-\frac{i}{n}\right)^{n-i}S_{n-i}^{(k-1)}<\frac{\sqrt{\pi}}{\Gamma(k/2)}\left(\frac{n}{2}\right)^{(k-1)/2} \tag{15}$$

$\Gamma(z)$ — гамма-функция, а $S_j^{(1)}=1$ для всех $j\geqslant1$. Поясним, что $S_n^{(k)}$ — сумма $p(\alpha^n|\Theta_0^m)$ для всех блоков α^n, содержащих k определенных букв алфавита A. Поскольку k различных букв можно выбрать C_m^k способами, отсюда сразу следует (14). Сравнительно просто выводится и рекуррентное выражение в (15). После этого неравенство в (15) доказывается индукцией по k.

Ограничение области значений θ любой $(m-1)$-мерной, не зависящей от n, областью $\Theta\subset\Theta_0^m$ уменьшает S_n не более чем в $c(\Theta)=\mathrm{const}$ раз и не изменяет асимптотическое поведение ρ_n, но при малых n возможны значительные различия (например, если Θ определяется неравенством $\theta_1\geqslant\ldots$ $\ldots\geqslant\theta_m$, то

$$S_1(\Theta)=1^{-1}+2^{-1}+\ldots+m^{-1}<\ln m+C+1/(2m),$$

где $C=0{,}577\ldots$ — постоянная Эйлера, в то время как $S_1(\Theta_0^m)=m$). Если же Θ — объединение областей размерности $z\leqslant m_0-1$, где $m_0<m$ (например, любой источник использует не более m_0 букв из A), то главный член избыточности пропорционален m_0-1, а не $m-1$.

Пример 2. Если Θ_P — множество источников без памяти с пуассоновским распределением $p(a|\theta)=\theta^a\exp\{-\theta\log e\}/a!$ вероятностей появления букв a счетного алфавита $A=\{0, 1, 2, \ldots\}$ и $0\leqslant\theta\leqslant\theta_0$, то

$$S_n(\Theta_P)=\sum_{i=0}^{i_0}\frac{e^{-i}i^i}{i!}+\sum_{i=i_0+1}^{\infty}\frac{e^{-n\theta_0}(n\theta_0)^i}{i!}=\sqrt{\frac{2n\theta_0}{\pi}}+c(n\theta_0),$$

где i_0 — целая часть $n\theta_0$, $0{,}82<c(n\theta_0)\leqslant 1$.

Для оценки $\rho_n(\Omega)$ и $S_n(\Omega)$ можно использовать не только прямые вычисления, но и неравенства. Например, $S_n(\Omega')\leqslant S_n(\Omega)$, если $\Omega'\subset\Omega$. Так же просты и другие оценки.

Следствие 1. *Если Ω — объединение подмножеств Ω_i, $i=\overline{1, M}$, то*

$$0\leqslant\rho_n(\Omega)-\max\{\rho_n(\Omega_i),\ 1\leqslant i\leqslant M\}<n^{-1}(1+\log M). \tag{16}$$

Справедливость этого утверждения следует из неравенства

$$p(\alpha^n|\Omega)=\max_{1\leqslant i\leqslant M}p(\alpha^n|\Omega_i)\leqslant\sum_{i=1}^{M}p(\alpha^n|\Omega_i)\leqslant M\max_{1\leqslant i\leqslant M}p(\alpha^n|\Omega_i)$$

и такого же неравенства для сумм $S_n(\Omega)$ и $S_n(\Omega_i)$. В частности, если каждое Ω_i содержит единственный элемент ω_i, то

$$L[\alpha^n|\varphi_n^*(\Omega)]<\min\{-\log p(\alpha^n|\omega_i),\ 1\leqslant i\leqslant M\}+\log S_n(\Omega)+1,$$

где $S_n(\Omega)\leqslant M$ и $\rho_n(\varphi_n^*, \Omega)<n^{-1}(1+\log M)$ (подобная конструкция для M источников упоминалась в [8]).

Пусть для любой пары распределений $Q_n=\{q(\alpha^n)\}$ и $W_n=\{w(\alpha^n)\}$ на A^n

$$R(\alpha^n|Q_n, W_n)=q(\alpha^n)/w(\alpha^n), \tag{17}$$

а если $w(\alpha^n)=q(\alpha^n)=0$, то $R(\alpha^n|Q_n, W_n)$ принимается равным 1. Минимум этого отношения по всем $\alpha^n\in A^n$ не превышает 1, а максимум — не меньше 1. Поэтому из (12) следует, что для произвольного W_n

$$\min[p(\alpha^n|\Omega)/w(\alpha^n)]\leqslant S_n(\Omega)\leqslant\max[p(\alpha^n|\Omega)/w(\alpha^n)], \tag{18}$$

где минимум и максимум определяется на всех $\alpha^n\in A^n$. Это неравенство будет использовано ниже.

МВ-коды, с точностью до $1/n$ минимизирующие максимальную избыточность кодирования отдельных сообщений, нередко асимптотически оптимальны и по критериям средней избыточности $r_n(\varphi_n, \Omega)$ и $r_n^*(\varphi_n, \Omega)$. Для ряда важных случаев последнее утверждение доказано в [9—11], а для одного класса компактных параметрических множеств — в [12] (предложенные в [12] коды практически совпадают с МВ-кодами для соответствующих множеств). В то же время для определения длин кодовых слов МВ-кода достаточно вычислить или оценить сверху значения $p(\alpha^n|\Omega)$ и $S_n(\Omega)$, что, как правило, значительно проще нахождения кода, минимизирующего $r_n(\varphi_n, \Omega)$ или $r_n^*(\varphi_n, \Omega)$.

§ 3. Последовательное кодирование

Проблема кодирования состоит в выборе как множества $L_n=\{L(\alpha^n|\varphi_n), \alpha^n\in A^n\}$, так и способа построения слов $\varphi(\alpha^n)$ длины $L(\alpha^n|\varphi_n)$, обладающих свойством разделимости. Один из подходов к решению последней задачи — применение кодов Элайеса или Гилберта — Мура в сочетании с принципом приближенных арифметических вычислений (см., например, [4—6]). Любая реализация такого кода определяется последовательностью «распределений» $W_k=\{w(\alpha^k),\ \alpha^k\in A^k,\ k=0,\ 1,\ 2,\ldots\}$, согласованных в том смысле, что

$$(19)\qquad \sum_{a\in A} w(\alpha^k a)\leqslant w(\alpha^k),\qquad \forall\alpha^k\in A^k$$

(таким образом, сумма $w(\alpha^k)$ по всем $\alpha^k\in A^k$ может быть меньше 1). Кодирование и декодирование выполняются последовательно, по мере поступления букв сообщения и символов кодового слова соответственно. «Текущая» длина кодового слова $L(\alpha^k|\varphi)$ (число символов, выведенных из кодера после первых k этапов кодирования) в первом приближении удовлетворяет неравенству

$$(20)\qquad L(\alpha^k|\varphi)+\log w(\alpha^k)\leqslant\Delta L_0,\ \Delta L_0=\text{const},$$

и не зависит от n. Поэтому в (20) и далее коды Элайеса и Гилберта — Мура обозначаются через φ без индекса длины блока.

Пусть из каких-либо соображений выбрано множество длин кодовых слов $L_n'=\{L_n'(\alpha^n),\ \alpha^n\in A^n\}$, удовлетворяющее неравенству Крафта со знаком равенства (последнее условие введено только для простоты). Положив $w(\alpha^n)$ равными

$$(21)\qquad q_n(\alpha^n)=\exp\{-L'(\alpha^n)\},$$

получим, согласно (20) и (21), что длины кодовых слов кода Элайеса или Гилберта — Мура удовлетворяют неравенству $L(\alpha^n|\varphi)-L'(\alpha^n)\leqslant\Delta L_0$. А для обеспечения согласованности (19) при всех $k<n$ достаточно взять $w(\alpha^k)=q_n(\alpha^k)$, где

$$(22)\qquad q_n(\alpha^k)=\sum q_n(\alpha^k\beta^{n-k})$$

— маргинальное распределение на первых k буквах относительно $Q_n=\{q_n(\alpha^n),\ \alpha^n\in A^n\}$ (суммирование выполняется по $\beta^{n-k}\in A^{n-k}$). Таким образом, одна из основных задач реализации блоковых кодов Элайеса или Гилберта — Мура — организация вычисления или хранения распределений $\{q_n(\alpha^k),\ \alpha^k\in A^k\}$, $\overline{k=1,\ n}$, или условных вероятностей $\vartheta_n(a|\alpha^k)=q_n(\alpha^k a)/q_n(\alpha^k)$.

Последовательный характер преобразований позволяет использовать коды Элайеса или Гилберта — Мура для кодирования полубесконечных последовательностей: тогда, как правило, важно, чтобы при каждом n текущая длина кодового слова $L(\alpha^n|\varphi)$ как можно меньше отличалась от длины $L'(\alpha^n)$ выбранного (для рассматриваемой задачи) кода. Иногда необходимо кодировать последовательности (точнее — подпоследовательности) заранее неизвестной длины (см. § 5). В обоих случаях задача реализации кодов Элайеса или Гилберта — Мура сводится к выбору распределения $W=\{W_n,\ n=1,2,\ldots\}$, согласованного в смысле (19) на всех сообщениях конечной длины и при любом n, как можно меньше отличающегося от Q_n, где $Q=\{Q_n,\ n=1,2,\ldots\}$ характеризует выбранный метод кодирования (см. (21)).

Если Q согласованно, то $W=Q$. Но Q далеко не всегда обладает этим свойством: его, например, не гарантирует «регулярность» кода, как это утверждается в [12] (код назван регулярным, если $L'(\alpha^k a)\geqslant L'(\alpha^k)$ для любых $\alpha^k\in A^k$, $a\in A$ и $k=0,\ 1,\ 2,\ldots$). А в противном случае различие W и Q неизбежно. Из (17), (20) и (21) следует, что при любом n

$$(23)\qquad L(\alpha^n|\varphi)-L'(\alpha^n)<\log R(\alpha^n|Q_n,\ W_n)+\Delta L_0.$$

Поэтому будем оценивать различие W и Q максимумом дополнительно вносимой избыточности (см. § 2)

$$\Delta\rho_n = n^{-1}\log\{\max R(\alpha^n|Q_n, W_n),\ \alpha^n\in A^n\}. \tag{24}$$

Определение искомого распределения W эквивалентно выбору условных вероятностей $\vartheta(a|\alpha^k)$. Любой функции $g(\alpha^k a)\geqslant 0$, определенной при некотором α^k и всех $a\in A$, можно поставить в соответствие условные вероятности

$$\vartheta(a|\alpha^k) = g(\alpha^k a)\Big[\sum_{\beta\in A} g(\alpha^k\beta)\Big]^{-1}, \tag{25}$$

а если $g(\alpha^k\beta)=0$ для всех $\beta\in A$, положим $\vartheta(a|\alpha^k)=1/m$. Например, представляет интерес выбор $g(\alpha^k a)=q_i(\alpha^k a)$, $i>k$ (см. (22)), при котором

$$\nu_i(a|\alpha^k) = q_i(\alpha^k a)\Big[\sum_{\beta\in A} q_i(\alpha^k\beta)\Big]^{-1}. \tag{26}$$

Подстановка $i=k+1$ минимизирует $\Delta\rho_{k+1}$ при любом $w(\alpha^k)$ (принцип локальной оптимизации). Можно также взять предел (26) при $i\to\infty$, если, конечно, он существует. Наконец, если

$$g(\alpha^k a) = \sup_{i>k}\vartheta_i(a|\alpha^k), \tag{27}$$

то

$$R(\alpha^n|Q_n, W_n) \leqslant \prod_{k=0}^{n-1}\Big[\sum_{\beta\in A}\sup\vartheta_i(\beta|\alpha^k)\Big].$$

В ряде случаев эти подходы обеспечивают хорошие результаты (см. § 4), но при некоторых Q заведомо плохи (см. § 5). Поэтому используем другой принцип выбора $\vartheta(a|\alpha^k)$.

Теорема 2. *Если $Q=\{Q_n, n=1,2,\ldots\}$ — последовательность произвольных распределений $Q_n=\{q_n(\alpha^n),\ \alpha^n\in A^n\}$, а $\lambda(n)$ — положительная функция натурального аргумента, для которой ряд $\lambda(1)+\lambda(2)+\ldots$ сходится к $\Lambda\leqslant 1$, то условные вероятности, задаваемые выбором*

$$g(\alpha^T) = \sum_{i=T}^{\infty}\lambda(i)\,q_i(\alpha^T), \qquad \alpha^T\in A^T, \qquad T=1,2,\ldots, \tag{28}$$

определяют согласованное распределение вероятностей W, при любом n удовлетворяющее неравенству

$$n\Delta\rho_n \leqslant \log\{\max\,[q_n(\alpha^n)/g(\alpha^n),\ \alpha^n\in A^n]\} \leqslant -\log\lambda(n) \tag{29}$$

(так как $\lambda(i)>0$ и $0\leqslant q_i(\alpha^T)\leqslant 1$, $i\geqslant T$, то из сходимости ряда $\lambda(1)+\lambda(2)+\ldots$ следует сходимость ряда в (28), и функция $g(\alpha^T)$ определена).

Доказательство. Поскольку сумма $q_i(\alpha^k\beta)$ по $\beta\in A$ равна $q_i(\alpha^k)$ (см. (22)), из (28) следует, что

$$\sum_{\beta\in A} g(\alpha^k\beta) = \sum_{i=k+1}^{\infty}\lambda(i)\,q_i(\alpha^k) = g(\alpha^k) - \lambda(k)\,q_k(\alpha^k), \qquad k>0,$$

а при $k=0$ получаем Λ. Если $q_n(\alpha^n)>0$, то при любом $k=\overline{1,\ n-1}$ эта сумма больше нуля и согласно (25)

$$w(\alpha^n) = \prod_{k=0}^{n-1}\vartheta(\alpha_{k+1}|\alpha^k) = \frac{g(\alpha^n)}{\Lambda}\prod_{k=1}^{n-1}\frac{g(\alpha^k)}{g(\alpha^k)-\lambda(k)\,q_k(\alpha^k)} \geqslant g(\alpha^n). \tag{30}$$

В противном случае $R(\alpha^n|Q_n, W_n)\leqslant 1$ (см. (17) и (25)). Из (24) и этих оценок сразу следует первое неравенство в (29). Второе неравенство тривиально следует из первого и (28). Теорема доказана.

Замечание 1. Нетрудно убедиться, что распределение вероятностей

$$\widetilde{w}(\alpha^n)=\left[\sum_{i=1}^{n-1} m^{i-n}\lambda(i)\,q_i(\alpha^i)+g(\alpha^n)\right]\Lambda^{-1}, \qquad \alpha^n\in A^n, \qquad n=1,2,\ldots$$

также согласовано и удовлетворяет (29). Но при согласованных Q_n, $n=1, 2, \ldots$, оно не совпадает с Q в отличие от (30). А замена суммы ряда (28) какой-либо оценкой может нарушить согласованность этого распределения, но не (30).

Следуя [13], введем

$$\lambda_0(x)=\left[\prod_{\tau=0}^{\tau(x)-1}\ln^\tau x\right]^{-1} \tag{31}$$

и

$$\lambda^*(x)=\lambda_0(x)\zeta[\tau(x)], \tag{32}$$

где $\ln^0 x=x$, $\ln^\tau x$, $\tau>0$, — τ-кратный натуральный логарифм x; $\tau(x)$ — целочисленное решение неравенства $0\leqslant\ln^\tau x<1$, $x\geqslant 1$; $\zeta(\tau)$ — положительная невозрастающая функция натурального аргумента, для которой

$$1{,}5\zeta(1)+1{,}082\zeta(2)+\zeta(3)+\zeta(4)+\ldots\leqslant 1.$$

Тогда $\lambda^*(n)$ удовлетворяет условиям теоремы 2 и

$$n\Delta\rho_n\leqslant-\log\lambda^*(n)=\log e\sum_{\tau=1}^{\tau(n)}\ln^\tau n-\log\zeta[\tau(n)]. \tag{33}$$

Дальнейшее уточнение оценок $\Delta\rho_n$ связано в первую очередь с конкретизацией последовательности Q, позволяющей улучшить оценки снизу для $q_i(\alpha^n)$, $i>n$, и соответственно $g(\alpha^n)$. Один из примеров такого уточнения приведен в § 4.

От одной последовательности Q несогласованных в общем случае распределений $\{Q_n, n=1, 2, \ldots\}$ легко перейти к конечному или счетному множеству $\{Q(j), j\in J\}$ таких последовательностей $Q(j)=\{Q_n^{(j)}, n=1, 2, \ldots\}$, где $Q_n^{(j)}=\{q_n^{(j)}(\alpha^n), \alpha^n\in A^n\}$. Если каждой $Q(j)$ поставлено в соответствие согласованное распределение $W(j)=\{w_j(\alpha^n), \alpha^n\in A^n, n=1, 2, \ldots\}$, вносящее дополнительную избыточность не более $\Delta\rho_n(j)$, $n=1, 2, \ldots$ то очевидно, что для любого распределения вероятностей $\{\xi(j)\}$ на J

$$w(\alpha^n)=\sum_{j\in J}\xi(j)\,w_j(\alpha^n)\geqslant\sup_{j\in J}\{\xi(j)\,w_j(\alpha^n)\} \tag{34}$$

— согласованное распределение, для которого

$$\log R(\alpha^n|Q_n^{(j)}, W_n)\leqslant n\Delta\rho_n(j)-\log\xi(j) \tag{35}$$

при всех $j\in J$, $\alpha^n\in A^n$ и $n=1, 2, \ldots$ В качестве $W(j)$ можно, например, использовать распределения вида (30).

§ 4. Универсальное последовательное кодирование

Рассмотрим последовательное кодирование источников с неизвестной статистикой. Как и в § 2, используем критерий максимума индивидуальной избыточности (5) для заданного множества источников Ω. В этом случае наилучшими являются распределения $Q_n=Q_n(\Omega)$, определяемые (12), и можно несколько конкретизировать результаты § 3.

Прежде всего, в соответствии с (21), (12) и (4)

$$(36)\qquad q_i(\alpha^n|\Omega)=[S_i(\Omega)]^{-1}\sum p(\alpha^n\beta^{i-n}|\Omega)\geqslant[S_i(\Omega)]^{-1}p(\alpha^n|\Omega)$$

(суммирование выполняется по $\beta^{i-n}\in A^{i-n}$). Кроме того, $S_i(\Omega)$ — неубывающая функция i для любого Ω и всегда можно ввести неубывающую функцию вещественного аргумента $S(x|\Omega)$, при натуральных $x=i$, равную $S_i(\Omega)$. Тогда для любого вещественного $X>n$ и $\lambda(i)=\lambda^*(i)$.

$$(37)\qquad g(\alpha^n)\geqslant p(\alpha^n|\Omega)\sum_{i=n}^{\infty}\frac{\lambda^*(i)}{S_i(\Omega)}>p(\alpha^n|\Omega)\zeta[\tau(X)]\int_n^X\frac{\lambda_0(x)\,dx}{S(x|\Omega)}$$

(интеграл можно оценивать снизу, заменяя, например, $S(x|\Omega)$ на $S(X|\Omega)$). Максимизируя правую часть этого выражения по X, во многих случаях удается усилить оценку (33).

Следствие 2. *Если для некоторого натурального b, вещественного положительного c, всех $n\geqslant n_0$, где $\tau(n_0)\geqslant b$, и всех $N>n$*

$$(38)\qquad \frac{S_N(\Omega)}{S_n(\Omega)}\leqslant\left(\frac{\ln^{b-1}N}{\ln^{b-1}n}\right)^c,$$

то для согласованного распределения, определяемого (28),

$$(39)\qquad n\Delta\rho_n<-\log\lambda_0(\ln^b n+1)-\log\zeta[\tau(n)+1]+\log(1+c),\quad n\geqslant n_0,$$

и если одновременно

$$(40)\qquad \lim_{n\to\infty}\frac{\ln^{b+1}n}{\ln S_n(\Omega)}=0,$$

то это распределение асимптотически оптимально, т. е.

$$(41)\qquad \Delta\rho_n(\Omega)=o[\rho_n(\Omega)],\ n\to\infty.$$

Доказательство. Существует функция $S(x|\Omega)$, удовлетворяющая (38) при замене N на x. Используя это неравенство в (37), оценим снизу полученный интеграл:

$$(42)\qquad \int_n^X\frac{\lambda_0(x)\,dx}{(\ln^{b-1}x)^c}=\int_{y_1}^{y_2}\frac{\lambda_0(z)\,dz}{z^c}>\lambda_0(\ln y_2)\int_{y_1}^{y_2}\frac{dz}{z^{1+c}},$$

где $y_1=\ln^{b-1}n$, $y_2=\ln^{b-1}X$. Сначала проведена последовательная замена переменных $x_j=\ln x_{j-1}$, $j=\overline{1,\ b-1}$ ($x_0=x$, $x_{b-1}=z$), при которой $\lambda_0(x_j)\,dx_j=\lambda_0(x_{j-1})\,dx_{j-1}$, а неравенство получено с помощью соотношения $z\lambda_0(z)=\lambda_0(\ln z)\geqslant\lambda_0(\ln y_2)$, $z\leqslant y_2$.

Выбирая X из условия

$$\ln^b X=\ln^b n+1$$

так, что $\tau(X)=\tau(n)+1$, и учитывая, что

$$c<(1+c)(1-e^{-c}),\quad c>0,$$

из (12), (37) и (42) получаем, что при $\lambda(i)=\lambda^*(i)$

$$g(\alpha^n)>q_n(\alpha^n|\Omega)\lambda_0(\ln^b n+1)\zeta[\tau(n)+1](1+c)^{-1}.$$

Подстановка этой оценки в (29) дает (39). При $n\to\infty$ правая часть (39) равна $O(\ln^{b+1}n)$ и из (40) сразу получаем (41). Следствие доказано.

Использованное при доказательстве значение X не оптимально, но нетрудно показать, что при любом выборе X можно уменьшить оценку сверху для $n\Delta\rho_n$ менее чем на $\log(1+c)$.

Замечание 2. Условие (38) ограничивает как величину $S_N(\Omega)$, так и скорость ее роста с увеличением N. Если же ограничивать лишь $S_N(\Omega)$, то оценка будет хуже. Например, если известно только, что

$$(43)\qquad c_2 \ln^b N \leqslant \ln S_N(\Omega) \leqslant c_1 \ln^b N, \quad 0 < c_2 \leqslant c_1 < \infty,$$

то к правой части (39) нужно добавить слагаемое $(c_1 - c_2)\ln^b N$, и для выполнения (41) необходимо и достаточно существование предела c отношения $[\ln S_N(\Omega)]/\ln^b N$ при $N\to\infty$. А для выполнения (39) нужно еще, чтобы это отношение стремилось к своему пределу сверху.

Даже при полностью известной $S(x|\Omega)$ погрешность оценок вида (37) может быть велика из-за использования неравенства (36). Но нахождение более точных оценок $q_i(\alpha^n|\Omega)$, $i>n$, затруднительно даже в простейших случаях. В то же время конкретный выбор условных вероятностей и непосредственное сравнение получаемой $w(\alpha^n)$ с $q_n(\alpha^n|\Omega)$ нередко позволяют значительно улучшить оценку $\Delta\rho_n$.

Вернемся к множеству Θ_0^m всех m-ичных источников без памяти (см. § 2), для которого

$$(44)\qquad p(\alpha^n|\Theta_0^m) = n^{-n} \prod_{a\in A} (t_a)^{t_a},$$

где $t_a = t_a(\alpha^n)$ — число вхождений буквы $a\in A$ в α^n. Из (12), (26) и (44) получаем, что

$$(45)\qquad \vartheta_{k+1}(a|\alpha^k) = \chi[t_a(\alpha^k)]\left\{\sum_{\beta\in A}\chi[t_\beta(\alpha^k)]\right\}^{-1},$$

где

$$\chi(z) = (z+1)^{z+1} z^{-z} = e[z+0{,}5 - (24z+12)^{-1} + O(z^{-3})],\quad z\to\infty$$

и

$$(46)\qquad \vartheta^*(a|\alpha^k, m) = \lim_{i\to\infty} \vartheta_i(a|\alpha^k) = \frac{t_a(\alpha^k)+1/2}{k+m/2}.$$

Нетрудно показать малое отличие (45) и (46), когда все $t_a(\alpha^k)$ больше нуля. Кроме того, насколько можно судить, $\vartheta_i(a|\alpha^k)$ — монотонные функции i, $i>k$. Последнее косвенно свидетельствует о малой несогласованности распределений $Q_n(\Theta_0^m)$ при разных n и позволяет использовать любые $\vartheta_i(a|\alpha^k)$ или (27). Для простоты выкладок используем (46); в этом случае

$$(47)\qquad w(\alpha^n|\Theta_0^m) = \frac{\Gamma(m/2)}{\pi^{m/2}\Gamma(n+m/2)} \prod_{a\in A} \Gamma(t_a + 1/2)$$

(см. также [6, 8, 14]). Можно показать, что $t\ln t - \ln\Gamma(t+{}^1/_2)$ — выпуклая вниз функция при $t\geqslant 0$ $(0\ln 0 = 0)$. Поэтому максимум $R(\alpha^n|Q_n, W_n) = R(\alpha^n|\Theta_0^m)$ (см. (17)) достигается, когда α^n — n-кратное повторение одной буквы, минимум — когда все t_a равны n/m и в общем случае

$$(48)\qquad R(\alpha^n|\Theta_0^m) = c_m(\alpha^n)\frac{\sqrt{\pi}}{\Gamma(m/2)}\left(n+\frac{m-1}{3}\right)^{(m-1)/2} = c_m(\alpha^n) R_0(n,m).$$

Здесь функция $R_0(n,m)$ определена вторым знаком равенства,

$$(49)\qquad 2^{-(m-1)/2} \leqslant c_m(\alpha^n) \leqslant 1,\quad m\geqslant 1,\ n\geqslant 1,$$

и знаки равенства в (49) имеют место только при $m=1$ (оценка вида (48) была получена в [8]). Из (3), (5), (18), (24), (48) и (49) следует, что при любом $n\geqslant 1$ последовательное кодирование φ с помощью условных вероятностей (46) обеспечивает

$$(50)\qquad \rho_n(\varphi, \Theta_0^m) < \frac{m-1}{2n}\log\left(n+\frac{m-1}{3}\right)+$$
$$+\frac{1}{n}\left[\frac{1}{2}\log\pi-\log\Gamma\left(\frac{m}{2}\right)+\Delta L_0\right]$$

(ΔL_0 введено в § 3), и для этого кодирования

$$(51)\qquad \Delta\rho_n \leqslant \frac{m-1}{2n}\log 2$$

(последняя оценка намного точнее (39) при $b=1$). Согласно (49), максимальная избыточность такого последовательного кодирования различных $\alpha^n \in A^n$ отличается менее чем на правую часть (51).

Заметим, что при использовании условных вероятностей (45) $\Delta\rho_n$ приблизительно в 1,5 раза меньше правой части (51), причем максимум индивидуальной избыточности, определяющий $\Delta\rho_n$, достигается на последовательностях сегментов длины m, каждый из которых содержит все буквы алфавита A. А при n-кратном повторении одной буквы избыточность приблизительно в $e/2$ раз меньше правой части (50).

§ 5. Марковские источники

Марковский источник ω имеет множество $U=\{u_v,\ v=\overline{1,\mu}\}$ состояний, в состоянии u_v с вероятностью $\theta(a|u_v, \omega)$ воспроизводит букву a алфавита $A_v \subset A$ из m_v букв, $1 \leqslant m_v \leqslant m$, и переходит в состояние $f(u_v, a) \in U$. Функции $f(u, a)$, определенной на $\nu=m_1+\ldots+m_\mu$ парах (u, a), $\mu \leqslant \nu \leqslant \leqslant m\mu$, соответствует направленный граф $\mathscr{F}$ с μ вершинами и ν дугами, а последовательности α^k, воспроизводимой из начального состояния u,— путь $\mathscr{P}(\alpha^k, u)$ в этом графе и μ подпоследовательностей независимых символов $\alpha_v^k(u)$, $v=\overline{1,\mu}$, порождаемых в каждом состоянии. Поэтому вероятность появления α^n на выходе источника ω полностью определяется μ распределениями вероятностей $\{\theta(a|u_v, \omega),\ a \in A_v\}$, $v=\overline{1,\mu}$, и распределением вероятностей начального состояния $\{\theta_0(u),\ u \in U\}$.

Рассмотрим универсальное последовательное кодирование множества F всех марковских источников с заданным графом $\mathscr{F}$. При этом используем некоторые дополнительные характеристики графа $\mathscr{F}$.

Дуги (переходы), по которым ни один путь $\mathscr{P}(\alpha^n, u)$ не проходит более 1 раза, назовем однократными, а остальные — многократными. Алфавит v-го состояния разделим на подмножества A_v' и A_v'' из $\overline{m}_v$ и $\Delta m_v = m_v - \overline{m}_v$ букв, определяющих многократные и однократные переходы из u_v соответственно, а все состояния источника — на три класса (типа) U_1, U_2 и U_3 в зависимости от значения Δm_v: 1) $\Delta m_v = 0$; 2) $0 < < \Delta m_v < m_v$; 3) $\Delta m_v = m_v$.

Пусть $\mu_1(\alpha^n, u)$ и $\mu_2(\alpha^n, u)$ — число вершин первого и второго типа, через которые проходит путь $\mathscr{P}(\alpha^n, u)$;

$$(52)\qquad \sigma(\alpha^n, u) = \sum (\overline{m}_v - 1),$$

где суммирование выполняется по всем этим вершинам; $\varkappa(\alpha^n, u)$ — общее число однократных переходов из вершин второго типа на пути $\mathscr{P}(\alpha^n, u)$ (согласно определению, из любого состояния возможен только один однократный переход и $\varkappa(\alpha^n, u) \leqslant \mu_2(\alpha^n, u)$); $\mathscr{M}(\alpha^n, u)$ — произведение Δm_v по всем вершинам пути, из которых выполнялись однократные переходы (если таких вершин нет, считаем $\mathscr{M}(\alpha^n, u)$ равным 1). Для $n \geqslant m\nu$ обозначим

$$(53)\qquad \sigma(\mathscr{F}) = \max \sigma(\alpha^n, u),$$

где максимум берется по всем $\mathscr{P}(\alpha^n, u)$; $\varkappa(\mathscr{F})$ — максимум $\varkappa(\alpha^n, u)$ по всем $\mathscr{P}(\alpha^n, u)$, на которых достигается $\sigma(\mathscr{F})$. Аналогичным образом последовательно определим $\mu_2(\mathscr{F})$ и $\mathscr{M}(\mathscr{F})$.

Введем функцию натурального аргумента

$$\psi(n)=\lambda^*(n)\,[\lambda^*(n)+\lambda^*(n+1)+\ldots]^{-1}\leqslant[e(n+m/2)]^{-1}, \tag{54}$$

где неравенство эквивалентно очень слабому ограничению скорости убывания функции $\zeta(\tau)$ в (32) (при $\tau\to\infty$ вполне достаточно выполнения условия $\zeta(\tau)\geqslant\exp(-c\tau)$ при любом $c<\infty$). Пусть, кроме того, $j(\alpha^k, u)$ — индекс (номер) конечного состояния пути $\mathscr{P}(\alpha^k, u)$. Тогда справедливо следующее утверждение.

Теорема 3. *Если задан граф $\mathscr{F}$ марковского источника, и $w(\alpha^n|u)$ — согласованное распределение вероятностей для начального состояния $u\in U$, определяемое условными вероятностями*

$$\vartheta(a\,|\,\alpha^k, u)=\begin{cases}\vartheta^*[a\,|\,\alpha_j^k(u), m_j], & u_j\in U_1,\ u_j\in U_3\\ [1-\psi(l_j+1)]\,\vartheta^*[a\,|\,\alpha_j^k(u), \overline{m}_j], & u_j\in U_2,\ a\in A_j',\\ \psi(l_j+1)/\Delta m_j & u_j\in U_2,\ a\in A_j'',\end{cases} \tag{55}$$

где $j=j(\alpha^k, u)$, $l_j=l_j(\alpha^k, u)$ — длина последовательности $\alpha_i^k(u)$, то последовательное кодирование φ на основе согласованного распределения вероятностей

$$w(\alpha^n)=\mu^{-1}[w(\alpha^n|u_1)+\ldots+w(\alpha^n|u_\mu)] \tag{56}$$

обеспечивает для множества F источников и всех n, начиная с некоторого

$$\frac{\rho_n(\varphi, F)}{\log e}<\frac{\sigma(\mathscr{F})}{2n}\ln\left[\frac{m-1}{\sigma(\mathscr{F})}n+\frac{m-1}{3}\right]+\frac{\varkappa(\mathscr{F})}{n}\ln\left[\frac{1}{2\lambda_0(\ln n)}\right]+ \tag{57}$$

$$+\frac{\mu_2(\mathscr{F})}{n}\ln\frac{1}{\zeta[\tau(n)]}+\frac{1}{n}\left\{\frac{\sigma(\mathscr{F})}{m-1}\ln\left[\frac{\sqrt{\pi}}{\Gamma(m/2)}\right]+\right.$$

$$\left.+\ln[\mu\mathscr{M}(\mathscr{F})]+\frac{\Delta L_0}{\log e}\right\}$$

(функции $\vartheta^*(\cdot)$, $\lambda^*(\cdot)$ и $\zeta(\cdot)$ определены в (46) и (32)).

Доказательство. Для любого начального состояния $u\in U$ условные вероятности (55) зависят только от подпоследовательности букв, уже воспроизведенных в состоянии u_j и $w(\alpha^n|u)$ — произведение μ сомножителей. Аналогичное разложение на сомножители имеет место для $p(\alpha^n|F, u)$ — максимальной вероятности α^n при заданных $\mathscr{F}$ и u. Поэтому отношение $p(\alpha^n|F, u)$ к $w(\alpha^n|u)$ — произведение таких отношений для отдельных состояний.

Согласно (55), для состояния первого типа это отношение равно $R(\alpha^{l_v}, \Theta_0^{m_v})$ (см. (48)), где $l_v=l_v(\alpha^n, u)$, для состояния третьего типа — $1/\Delta m_v$ (в такое состояние можно попасть только 1 раз) и наконец, для состояния второго типа

$$R(\alpha^{l-1}|\Theta_0^{\overline{m}_v})\,\frac{(l-1)^{l-1}}{l^l}\left\{\prod_{k=1}^{l}[1-\psi(k)]\right\}^{-1}\times$$

$$\times\left[\Delta m_v\frac{1-\psi(l)}{\psi(l)}\right]^{\delta}\left[\frac{(t+1)^{t+1}}{t^t}\,\frac{l-1+\overline{m}_v/2}{t+1/2}\right]^{1-\delta}$$

где $l=l_v(\alpha^n, u)$; $\delta=0$, если из состояния u_v не выполнялся однократный переход, $\delta=1$ в противном случае; $t=t_a(\alpha_v^n)$, где $a=\alpha_l\in A_v'$ (для $\delta=0$). При получении этого выражения учитывалось, что из любого состояния возможен только один однократный переход и только при воспроизведении последней буквы $\alpha_v^n(u)$ (после такого перехода возвращение в u_v невозможно).

Из неравенства в (54) следует, что максимум последнего выражения достигается при $\delta=1$ и согласно определению (54) может быть представлен в виде

$$R(\alpha^{l-1}|\Theta_0^{\bar{m}_v})\frac{(l-1)^{l-1}\Delta m_v}{l^l\lambda^*(l)}\leqslant R_0(l,\bar{m}_v)\frac{\Delta m_v}{2l\lambda^*(l)},\qquad l>1$$

(случай $l=1$ не представляет интереса). Учитывая, что $l\lambda^*(l)\geqslant n\lambda^*(n)=$ $=\zeta[\tau(n)]\lambda_0(\ln n)$, для произведения всех μ отношений получаем

$$\text{(58)}\qquad \frac{p(\alpha^n|F,u)}{w(\alpha^n|n)}<[\Pi R_0(l_v,\bar{m}_v)][2\lambda_0(\ln n)]^{-\varkappa}[\zeta(\tau(n))]^{-\mu_2}\mathscr{M}(\alpha^n,u),$$

где произведение берется по всем состояниям первого и второго типов, через которые проходит путь $\mathscr{P}(\alpha^n,u)$, $\varkappa=\varkappa(\alpha^n,u)$ и $\mu_2=\mu_2(\alpha^n,u)$. Используя определение $R_0(n,m)$ в (48), нетрудно показать, что

$$\text{(59)}\qquad \Pi R_0(l_v,\bar{m}_v)\leqslant\Pi R_0\left[\frac{\bar{m}_v-1}{\sigma}n,\bar{m}_v\right]\leqslant$$
$$\leqslant\left[\frac{\sqrt{\pi}}{\Gamma(m/2)}\right]^{\sigma/(m-1)}\left(\frac{m-1}{\sigma}n+\frac{m-1}{3}\right)^{\sigma/2},$$

где $\sigma=\sigma(\alpha^n,u)$, а второе неравенство следует из выпуклости вниз функции $(z-1)\ln(z-1)-2\ln\Gamma(z/2)$ при $z\geqslant1$.

Очевидно, что $p(\alpha^n|F)=\max\{p(\alpha^n|F,u),\ u\in U\}$ и согласно (56)

$$\text{(60)}\qquad \frac{p(\alpha^n|F)}{w(\alpha^n)}\leqslant\mu\max_{u\in U}\frac{p(\alpha^n|F,u)}{w(\alpha^n|u)}.$$

Сомножители в правой части (58) расположены в порядке убывания скорости их роста по n. Поэтому из (58)–(60), (3) и (5) следует справедливость (57), начиная с некоторого n. Теорема доказана.

Замечание 3. В (55) можно использовать любую функцию $\psi(l)$, $0<\psi(l)<1$, но нетрудно убедиться, что выбор (54) минимизирует верхнюю оценку избыточности при больших n.

Замечание 4. Если граф $\mathscr{F}$ содержит однократные дуги, то при малых n максимум отношения (58) может достигаться на пути с $\sigma(\alpha^n,u)<\sigma(\mathscr{F})$ и $\varkappa(\alpha^n,u)>\varkappa(\mathscr{F})$ — именно поэтому неравенство (57) начинает выполняться только для n, больших некоторого. Если же однократных дуг нет, оценка справедлива для $n\geqslant1$.

Чтобы определить зависящие от $\mathscr{F}$ коэффициенты в (57), следует разложить $\mathscr{F}$ на сильно связные графы и состояния третьего типа, соединенные однократными дугами (граф называется сильно связным, если из любого его состояния можно попасть в любое состояние за конечное число шагов). Но можно подставить в (57) и любые верхние оценки этих коэффициентов. Например, коэффициент $\sigma(\mathscr{F})$, играющий наиболее важное значение, удовлетворяет неравенству

$$\text{(61)}\qquad \sigma(\mathscr{F})\leqslant(\nu-\Delta\nu)-(\mu-\Delta\mu)\leqslant\nu-\mu,$$

где $\Delta\nu$ — число однократных дуг в $\mathscr{F}$, а $\Delta\mu$ — число состояний третьего типа. Действительно, для всех состояний первого и второго типов $\bar{m}_v-1\geqslant0$ и для оценки $\sigma(\mathscr{F})$ сверху достаточно суммировать в (52) по всем таким состояниям. Отсюда получаем первое неравенство. А поскольку $\Delta\nu\geqslant\Delta\mu$, очевидно и второе неравенство.

Допуская замену $\sigma(\mathcal{F})$ на $\nu-\mu$, можно существенно упростить оценку (57).

Следствие 3. *Если при последовательном кодировании во всех состояниях марковского источника используются условные вероятности* $\vartheta^*[a|\alpha_j^k(u), m_j]$, *то при любом* $n\geqslant 1$

$$\frac{\rho_n(\varphi,F)}{\log e}<\frac{\nu-\mu}{2n}\ln\left(\frac{m-1}{\nu-\mu}n+\frac{m-1}{3}\right)+ \tag{62}$$

$$+\frac{1}{n}\left\{\frac{\nu-\mu}{m-1}\ln\left[\frac{\sqrt{\pi}}{\Gamma(m/2)}\right]+\ln\mu+\frac{\Delta L_0}{\log e}\right\}.$$

Действительно, при унифицированном выборе переходных вероятностей в правую часть (58) входит только первый сомножитель, произведение берется по всем $u_v \in U$, а $\overline{m}_v$ заменены на m_v. Отметим, что оценка (62) справедлива для любого графа с заданными ν и μ.

Если $\mathcal{F}$ — совокупность любого числа изолированных сильно связных подграфов, то вместо $\nu-\mu$ можно подставить максимальную разность числа дуг и вершин в отдельном подграфе. А если $\mathcal{F}$ — сильно связный граф, то (62) совпадает с (57). Так, множеству m-ичных марковских цепей связности s (когда состояние источника определяется s последними буквами на его выходе) соответствует сильно связный граф с m^{s+1} дугами и m^s вершинами. Подставив эти значения в (62), получим, что при $\Delta L_0<1$ (что в принципе возможно) эта оценка сильнее полученной в [15] для *средней* избыточности $r_n^*(\varphi_n, F)$.

При заданном начальном состоянии u и $i=k+1$ переходные вероятности (26) определяются (45) с заменой α^k на $\alpha_j^k(u)$, A на A_j, где $j=j(\alpha^k, u)$, и хорошо аппроксимируются $\vartheta^*[a|\alpha_j^k(u), m_j]$ для состояний любого типа. Из сравнения (57) и (62) следует, что для состояний второго типа такой выбор неоптимален. А предел выражения (26) при $i\to\infty$ вообще невозможно использовать — нетрудно показать, что для однократных переходов он равен нулю. Последнее означает, что (27) мало отличается от $\vartheta_{k+1}(a|\alpha^k)$ и в такой же степени неоптимальна. Таким образом, среди общих подходов к выбору условных вероятностей, обсуждавшихся в § 3, в рассматриваемом случае наиболее эффективен предложенный в теореме 2, и можно предположить, что (55) является некоторой аппроксимацией (28).

Для сильно связных графов $\mathcal{F}$ главный член нижней оценки $r_n(F)$ совпадает с первым членом в (62) (см. [11]) и согласно (7) предложенное последовательное кодирование асимптотически оптимально. Более того, есть все основания полагать, что в этом случае $\Delta\rho_n=O(1/n)$ (для марковских цепей связности s это следует из (62) и нижней оценки $r_n(F)$, полученной в [15]). В то же время из прямых вычислений $S_n(F)$ следует, что

$$n\rho_n(F)<0.5\sigma(\mathcal{F})\log n+\varkappa(\mathcal{F})\log\log n+\log C(\mathcal{F}), \tag{63}$$

где $C(\mathcal{F})$ не зависит от n, и при $\varkappa(\mathcal{F})>0$ для рассмотренного метода

$$n\Delta\rho_n(F)\geqslant\varkappa(\mathcal{F})\ln^3 n. \tag{64}$$

Эта оценка представляется асимптотически точной (для доказательства нужно найти нижнюю оценку $\rho_n(F)$). Кроме того, учитывая достаточно общий вид переходных вероятностей (55) для $u_j \in U_2$ и оптимизацию по $\psi(l)$, естественно предположить, что при любом другом последовательном кодировании можно уменьшить правую часть (57) не более чем на $O(1/n)$.

Оценки (57) и (62) справедливы для произвольных распределений начального состояния и, таким образом, для нестационарных $p(\alpha^n|\omega)$ (при наличии однократных переходов невырожденные стационарные распределения вообще не существуют). Именно для того чтобы обеспечить равномерность полученных оценок по всем распределениям начального состояния и избежать необходимости поиска начального состояния, наиболее «удобного» для описания α^n, и было введено согласованное распределение (56). Этот же подход можно применить еще раз.

Следствие 4. *Если G — объединение M множества марковских источников F_i, $i=\overline{1, M}$, с различными графами, то при выборе*

(65) $$w^*(\alpha^n)=M^{-1}[w(\alpha^n|F_1)+\ldots+w(\alpha^n|F_M)],$$

где для любого F_i распределение $w(\alpha^n|F_i)$ определяется в соответствии с (56),

(66) $$\rho_n(\varphi, G)<\max\{\rho_n(\varphi, F_i),\ 1\leqslant i\leqslant M\}+n^{-1}(\log M+1).$$

Доказательство этого утверждения, аналогичного (16), очевидно. Использование (65) позволяет осуществлять универсальное последовательное кодирование для множества всех марковских источников с ограниченными μ (и ν), множества марковских цепей связности $s\leqslant s_0$ и т. д. Оценка (66) справедлива и в тех случаях, когда $M=M(n)$ и $G=G_n$ зависят от n. Поэтому можно расширять множество $G=G_n$ с ростом n, контролируя величину дополнительной избыточности $n^{-1}\log M(n)$ (например, она должна стремиться к нулю с ростом n, или не превышать первого слагаемого в (66), или быть асимптотически малой по сравнению с этим слагаемым и т. д.). Так, для множества G марковских цепей с $s\leqslant s_0(n)$ дополнительная избыточность не превышает $(\log n)/n$, если $s_0(n)=n$, и асимптотически мала по сравнению с главным членом, если $s_0(n)$ растет медленнее любой степени n.

Любому полному дереву «контекстов», введенных в [16], соответствует сильно связный граф марковского источника. Поэтому сказанное выше в полной мере распространяется на множество источников с заданным контекстом и на ограниченное множество контекстов. Например, общее число M различных контекстных деревьев с длиной ветвей не более D удовлетворяет неравенству

$$m^{D-1}\ln 2<\ln M<m^D\ln 2,\quad D>1,$$

и легко определить допустимую скорость увеличения $D=D(n)$ как функцию требуемой скорости уменьшения дополнительной избыточности.

§ 6. Заключение

Как уже отмечалось, во многих важных случаях максимальная избыточность кодирования отдельных сообщений асимптотики совпадает со средней и простые по конструкции коды максимальных вероятностей близки к оптимальным сразу по обоим критериям. Вместе с тем существуют и различия между этими подходами.

Например, известно существование универсальных (по критерию $r_n^*(\varphi_n, \Omega)$) кодов для любого множества стационарных источников, для каждого из которых $n^{-1}H_\omega(A^n)-H_\omega\leqslant\varepsilon(n)$, где $\varepsilon(n)$ стремится к нулю с ростом n. При этом источник приближенно описывается $s(n)$-связной марковской цепью, где $s(n)\to\infty$ при $n\to\infty$. В то же время неясно, можно ли при такой аппроксимации обеспечить стремление $\rho_n(\alpha^n|\varphi_n, \Omega)$ к нулю *для всех* α^n, а не с вероятностью единица (последнее очевидно).

Различие подходов проявляется также, когда, например, для одного из сообщений $p(\alpha^n|\omega)$ известно (одинаково) для всех $\omega\in\Omega$. По критерию $r_n^*(\varphi_n, \Omega)$ или $r_n(\varphi_n, \Omega)$ следует приписать этому блоку его известную вероятность, а по критерию $\rho_n(\varphi_n, \Omega)$ — величину $p(\alpha^n|\omega)/S_n(\Omega)<p(\alpha^n|\omega)$. Но можно несколько модифицировать МВ-метод, используя

(67) $$\tilde{p}(\alpha^n|\Omega)=p'(\alpha^n|\Omega)+\frac{1-S_{\min}^{(n)}(\Omega)}{S_n(\Omega)-S_{\min}^{(n)}(\Omega)}[p(\alpha^n|\Omega)-p'(\alpha^n|\Omega)],$$

где $p'(\alpha^n|\Omega)=\inf\{p(\alpha^n|\omega), \omega\in\Omega\}$, а $S^{(n)}_{\min}(\Omega)\leqslant 1$ — сумма этих вероятностей по всем $\alpha^n\in A^n$. Только если все $p'(\alpha^n|\Omega)=0$, определяемый этим распределением код совпадает с МВ-кодом.

Очевидно, что принцип максимальных вероятностей может быть использован для простого построения кодов, удовлетворяющих самым разным критериям эффективности. Более подробное изучение этого вопроса — тема отдельного рассмотрения.

Переход от блокового к последовательному кодированию позволяет использовать достоинства разработанных в последнее время методов ценой сравнительно небольшого увеличения избыточности.

ЛИТЕРАТУРА

1. *Штарьков Ю. М.* Кодирование сообщений конечной длины на выходе источника с неизвестной статистикой // Тр. V Всесоюз. конф. по теории кодирования и передачи информации. Тез. докл. Москва — Горький, 1972. Ч. I. С. 147—152.
2. *Штарьков Ю. М.* Применение метода максимального правдоподобия для кодирования источников с неизвестной статистикой // Тр. III Междунар. симпоз. по теории информации. Тез. докл. Москва — Таллин, 1973. Ч. II. С. 176—180.
3. *Shtar'kov Yu. M.* Coding of discrete sources with unknown statistics // Topics on Information Theory (Second Colloquium, Keszthely, 1975). Colloquia Mathematica Sosietatis Janos Bolyai. Amsterdam: Nort Holland, 1977. V. 16. P. 559—574.
4. *Rissanen J., Langdon G. G.* Universal Modeling and Coding // IEEE Trans. Inform. Theory. 1981. V. 27. № 1. P. 12—23.
5. *Jones C. B.* An Efficient Coding System for Long Source Sequences // IEEE Trans. Inform. Theory. 1981. V. 27. № 3. P. 280—291.
6. *Штарьков Ю. М.* Обобщенные коды Шеннона // Пробл. передачи информ. 1984. Т. 20. № 3. С. 3—16.
7. *Галлагер Г.* Теория информации и надежная связь. М.: Сов. радио, 1974.
8. *Davisson L. D., McElice R. J., Pursley M. B., Wallace M. S.* Efficient Universal Noiseless Source Codes // IEEE Trans. Inform. Theory. 1981. V. 27. № 3. P. 269—279.
9. *Кричевский Р. Е.* Лекции по теории информации. Новосибирск: Изд-во НГУ, 1970.
10. *Трофимов В. К.* Избыточность универсального кодирования произвольных марковских источников // Пробл. передачи информ. 1974. Т. 10. № 4. С. 16—24.
11. *Штарьков Ю. М.* Метод построения нижних границ избыточности универсального кодирования // Пробл. передачи информ. 1982. Т. 18. № 2. С. 3—11.
12. *Rissanen J.* Universal Coding, Information, Prediction and Estimation // IEEE Trans. Inform. Theory. 1984. V. 30. № 4. P. 629—636.
13. *Левенштейн В. И.* Об избыточности и замедлении разделимого кодирования натуральных чисел // Проблемы кибернетики. М.: Наука, 1968. Вып. 20. С. 173—179.
14. *Krichevsky R. E., Trofimov V. K.* Optimal sample-based encoding Markov sources // Proc. 3-rd Czech.— Sov.— Hung. Semin. Inform. Theory, Liblice, 1980. Liblice, 1980. P. 131—137.
15. *Davisson L. D.* Minimax Noiseless Universal Coding for Markov Sources // IEEE Trans. Inform. Theory. 1983. V. 29. № 2. P. 211—215.
16. *Rissanen J.* A Universal Data Compression System // IEEE Trans. Inform. Theory. 1983. V. 29. № 5. P. 656—664.

Поступила в редакцию
26.VI.1985