ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь—март 2004. Серия 1. Том 11, № 1, 52–78

УДК 519.714

АДДИТИВНАЯ СЛОЖНОСТЬ СЛОВ С ОГРАНИЧЕНИЯМИ НА СОСТАВ ПОДСЛОВ*)

В. Н. Потапов

Аддитивной сложностью слова называется длина кратчайшей схемы конкатенации, порождающей это слово. Для слов длины n, удовлетворяющих различным ограничениям на состав подслов, получены асимптотические (при $n \to \infty$) верхние оценки аддитивной сложности. Показано, что эти оценки неулучшаемы для наиболее сложных слов из рассматриваемых множеств.

Введение

Пусть A — конечный алфавит и $A^* = \bigcup\limits_{m=0}^{\infty} A^m$ — множество конечных слов в алфавите $A, |A| \geqslant 2$. Через wu будем обозначать конкатенацию слов $u, w \in A^*$. Символом λ обозначим "пустое" слово. *Схемой конкатенации* называется последовательность слов $\alpha = v(0), v(1), \ldots, v(m)$ такая, что каждое слово $v(i)$ является либо буквой из A, либо конкатенацией предыдущих слов $v(j_1)$ и $v(j_2)$, т. е. $v(i) = v(j_1)v(j_2)$, где $j_1, j_2 < i$. Говорят, что $v(0), v(1), \ldots, v(m)$ — схема конкатенации множества $M \subset A^*$, если все слова множества M содержатся в схеме конкатенации, т. е. $M \subseteq \{v(0), v(1), \ldots, v(m)\}$. *Аддитивной сложностью* множества M называется величина $l(M) = \min m$, где минимум берется по всем схемам конкатенации множества M. В работах [5], [8], [9] указанная величина называлась мультипликативной сложностью. Однако по предложению А. А. Евдокимова термин "мультипликативная" заменен на более подходящий термин "аддитивная". Ниже речь пойдет в основном об аддитивной сложности множеств, состоящих из одного слова $w \in A^*$, которую будем обозначать через $l(w)$.

*) Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проект 02-01-00939), Министерства образования (проект Е 02–6.0–250), Совета по грантам Президента РФ и государственной программы поддержки ведущих научных школ (проект НШ–313.2003.1) и СО РАН (интеграционный проект 119).

Пусть $B \subset A^*$. Определим величину $L(B)$ равенством

$$L(B) = \max_{w \in B} l(w).$$

В работе [5] установлена следующая асимптотика (при $n \to \infty$):

$$L(M(n-k_n, k_n)) \sim \log n + \frac{\log |M(n-k_n, k_n)|}{\log\log |M(n-k_n, k_n)|},$$

где $M(n-k_n, k_n)$ — множество двоичных слов длины n, в каждом из которых содержится k_n единиц. Здесь и далее $\log x = \log_2 x$. В настоящей работе установлены асимптотики (при $n \to \infty$) вида $L(B_n) \sim \dfrac{\log |B_n|}{\log\log |B_n|}$, где $B_n \subset A^n$ есть множество слов с одинаковым составом подслов фиксированной длины, или множество слов с запретами на подслова, или множество высоковероятных слов относительно некоторых вероятностных распределений. Некоторые из полученных асимптотик справедливы при определенных ограничениях снизу на скорость роста мощности множества B_n при $n \to \infty$.

Нижние оценки величины $L(B_n)$ для разных B_n получены мощностным методом и основываются на результатах статьи [5]. Как замечено в работе [6], следующее утверждение непосредственно вытекает из [6, теорема Д.2].

Утверждение 1. *Пусть $B_1, B_2, \ldots, B_n, \ldots$ — последовательность множеств таких, что $B_n \subset A^*$ и $h_n = \log |B_n| \to \infty$ при $n \to \infty$. Тогда*

$$L(B_n) \geqslant \frac{h_n}{\log h_n}\Big(1 + (1+o(1))\frac{\log h_n}{\log\log h_n}\Big).$$

Верхние оценки для величины $L(B_n)$ получены конструктивно. Как в работе [5], сначала определяются величины $l(M_\alpha)$ для некоторых вспомогательных множеств $M_\alpha \subset A^*$, а затем из элементов этих множеств строятся схемы конкатенации для слов $w \in B_n$. В отличие от работы [5] для построения множеств M_α используется вероятностный подход: на множестве слов A^* определяется вероятностное распределение и в множество M_α включаются слова с вероятностями выше фиксированной. Это позволяет упростить структуру вспомогательных множеств M_α, обобщить результаты статьи [5] и в некоторых случаях получить более точные верхние оценки. Результаты настоящей статьи анонсированы в [8], [9].

1. Марковские источники

Одним из наиболее важных инструментов, применяемых в настоящей статье, является понятие марковского источника и сопутствующий математический аппарат. *Марковским источником* S называется четверка (A, Ω, μ_S, p_S), состоящая из конечного алфавита $A = \{a_1, \ldots, a_{|A|}\}$, конечного множества состояний $\Omega = \{\sigma_1, \ldots, \sigma_{|\Omega|}\}$, функции $\mu_S : \Omega \times A \to \Omega$, задающей переходы из одного состояния в другое, и набора вероятностей $p_S(a|\sigma)$ букв $a \in A$ в состояниях $\sigma \in \Omega$, $0 \leqslant p_S(a|\sigma) \leqslant 1$. В другой терминологии марковский источник можно называть стохастическим конечным автоматом, а множество порождаемых им слов — стохастическим регулярным языком.

Функцию μ_S удобно представлять в виде ориентированного графа $G(\mu_S)$, вершинами которого являются состояния источника. Два состояния σ' и σ'' соединены дугой, помеченной буквой a, если $\mu_S(\sigma', a) = \sigma''$ и $p_S(a|\sigma') > 0$. Источник S называется *неразложимым*, если ориентированный граф $G(\mu_S)$ является сильно связаным. Неразложимый источник S называется *периодическим*, если в ориентированном графе $G(\mu_S)$ длины всех контуров кратны некоторому целому числу $d > 1$. Если $d = |\Omega|$, т. е. ориентированный граф $G(\mu_S)$ состоит из одного контура, то источник S будем называть *тривиальным*. В последнем случае вероятности всех букв равны 0 или 1.

Расширим область определения функции μ_S и вероятности p_S с алфавита A на множество слов A^* с помощью рекуррентных формул:

$$\mu_S(\sigma, a_{i_1} a_{i_2} \ldots a_{i_m}) = \mu_S(\mu_S(\sigma, a_{i_1}), a_{i_2} \ldots a_{i_m}), \tag{1}$$

$$p_S(a_{i_1} a_{i_2} \ldots a_{i_m}|\sigma) = p_S(a_{i_2} \ldots a_{i_m}|\mu_S(\sigma, a_{i_1}))p_S(a_{i_1}|\sigma). \tag{2}$$

Пусть слово $w \in A^*$ представлено в виде $w = w_1 w_2 \ldots w_m$. Тогда из равенств (1) и (2) получаем

$$p_S(w|\sigma) = p_S(w_1|\tau_1)p_S(w_2|\tau_2) \ldots p_S(w_m|\tau_m), \tag{3}$$

где $\tau_1 = \sigma$ и $\tau_j = \mu_S(\tau_{j-1}, w_{j-1})$.

По определению вероятности букв в произвольном состоянии $\sigma \in \Omega$ удовлетворяют равенству $\sum\limits_{a \in A} p_S(a|\sigma) = 1$. Тогда из рекуррентной формулы (3) для любого слова $w \in A^*$ и любого целого числа $m > 0$ по индукции можно получить равенства

$$\sum_{a \in A} p_S(wa|\sigma) = p_S(w|\sigma) \text{ и } \sum_{w \in A^m} p_S(w|\sigma) = 1. \tag{4}$$

Последовательность состояний марковского источника образует однородную марковскую цепь, если вероятности переходов задать равенствами $p(\mu_S(\sigma, a)|\sigma) = p_S(a|\sigma)$. Данные выше определения неразложимого и периодического марковских источников соответствуют определениям неразложимой и периодической марковских цепей для состояний источника (см., например, [2, гл. 11, § 1]). В дальнейшем это позволит воспользоваться некоторыми фактами из теории марковских цепей.

Источник S называется *марковским источником k-го порядка*, если $\Omega \subseteq A^k$ и для каждой $a \in A$ справедливо равенство $\mu_S(a_{i_1}a_{i_2}\dots a_{i_k}, a) = a_{i_2}\dots a_{i_k}a$. В этом случае вклад вероятности буквы в вероятность слова зависит только от k предыдущих букв. Однако остается неопределенность в вычислении вероятности k первых букв каждого слова. Чтобы избежать этого, в дальнейшем для слова $w \in A^*$ полагаем $p_S(w) = p_S(w|w')$, где $w' \in A^k$ — суффикс длины k слова w. Таким образом, для марковского источника S k-го порядка вероятность слова длины $n, n > k$, вычисляем по формуле

$$\begin{aligned} p_S(a_{i_1}a_{i_2}\dots a_{i_n}) &= p_S(a_{i_1}|a_{i_{n-k+1}}\dots a_{i_n}) \\ &\quad\times p_S(a_{i_2}|a_{i_{n-k+2}}\dots a_{i_n}a_{i_1})\dots p_S(a_{i_n}|a_{i_{n-k}}\dots a_{i_{n-1}}), \quad (5) \end{aligned}$$

согласующейся с формулой (3).

Граф $G(\mu_S)$, соответствующий марковскому источнику k-го порядка, является подграфом графа де Брейна G_k, т. е. вершинами графа $G(\mu_S)$ являются слова $w \in A^k$ и слова вида bw' и $w'a$, где $w' \in A^{k-1}$ и $a, b \in A$, соединены дугой, ориентированной из w' в $w'a$. Пусть $w = a_{i_1}a_{i_2}\dots a_{i_n} \in A^n$ и $\tau, \sigma \in \Omega$. Введем обозначения:

$$I_S(w|\sigma) = -\log p_S(w|\sigma) = -\sum_{j=1}^{n} \log p_S(a_{i_j}|\tau_j), \quad (6)$$

где $\tau_1 = \sigma$, $\tau_j = \mu_S(\tau_{j-1}, a_{i_{j-1}})$,

$$I_{S,\tau}(w|\sigma) = -\sum \log p_S(a_{i_j}|\tau), \quad (7)$$

а сумма берется по всем j таким, что $\tau_j = \tau$. Очевидно справедливо равенство

$$I_S(w|\sigma) = \sum_{\tau \in \Omega} I_{S,\tau}(w|\sigma). \quad (8)$$

Если S — марковский источник k-го порядка, то для слова $w \in A^*$ длины $|w| > k$ аналогично определяются величины $I_S(w)$ и $I_{S,\tau}(w)$, где $\tau \in A^k$.

Пусть $x = x_1x_2\dots x_n\dots$ — символьная последовательность в алфавите A и слово $x^n = x_1x_2\dots x_n$ — префикс (начало) длины n последовательности x. Для неразложимого непериодического марковского источника

S величина $I_S(x^n|\sigma)/n$ сходится (при $n \to \infty$) по вероятности к энтропии источника $H(S) > 0$ независимо от $\sigma \in \Omega$ (см., например, [3, теорема 3.5.3]). Тем более имеет место сходимость в среднем:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{w \in A^n} p_S(w|\sigma) I_S(w|\sigma) = H(S).$$

Здесь и далее подразумевается, что $0 \log 0 = 0$.

В дальнейшем, когда это не приводит к неоднозначности, индекс S в обозначениях p_S, μ_S и I_S будем опускать.

2. Оценки аддитивной сложности через вероятности слов

Множество $T \subset A^*$, удовлетворяющее свойству

$$\text{если } w \in T \text{ и } w = uv, \text{ то } u \in T, \tag{9}$$

обладает структурой дерева — здесь предками вершины-слова являются префиксы слова. Множество слов, обладающее свойством (9), будем называть *деревом*.

Дерево $T \subset A^*$ будем называть *полным* , если

$$\text{для любых } a, b \in A \text{ из } wa \in T \text{ следует } wb \in T. \tag{10}$$

Обозначим через $g(T)$ множество слов, соответствующих листьям (висячим вершинам) дерева T. Нам понадобится следующее

Утверждение 2. *Пусть S — марковский источник в алфавите A с множеством состояний Ω и T — дерево в алфавите A. Тогда*

(a) для любого $\sigma \in \Omega$ справедливо неравенство $\sum_{u \in g(T)} p(u|\sigma) \leqslant 1$;

(b) $l(T) = |T| - 1$;

(c) если T — полное дерево, то $|T| = |A|(|T| - |g(T)|) + 1$;

(d) если каждое слово $w \in T \backslash g(T)$ является префиксом не менее двух слов из T, то $|T| \leqslant 2|g(T)| - 1$.

Доказательство. Индукцией по числу вершин дерева нетрудно показать, что для полного дерева справедливы равенства

$$(|T| - 1) = (|g(T)| - 1)\frac{|A|}{|A| - 1} \text{ и } \sum_{u \in g(T)} p(u|\sigma) = 1.$$

Отсюда следуют (a) и (c).

Неравенства $l(T) \leqslant |T| - 1$ и $|T| \leqslant 2|g(T)| - 1$ также получаются индукцией по числу вершин в дереве, а неравенство $l(T) \geqslant |T| - 1$ следует из определения аддитивной сложности. Утверждение 2 доказано.

Пусть S — марковский источник в алфавите A с множеством состояний Ω, $\sigma \in \Omega$ и $q > 0$. Множество

$$M_\sigma(q) = \{u \in A^* \mid p(u|\sigma) \geqslant q\}$$

удовлетворяет условию (9), т. е. является деревом. Рассмотрим полное дерево

$$T_\sigma(q) = \{u \in A^* \mid u = wa, w \in M_\sigma(q), a \in A \cup \{\lambda\}\}.$$

Заметим, что $g(T_\sigma(q)) = T_\sigma(q) \backslash M_\sigma(q)$ и для каждого $u \in g(T_\sigma(q))$ справедливо неравенство

$$p(u|\sigma) < q. \tag{11}$$

Введем обозначение

$$\varrho_S = 1 - \max_{\sigma \in \Omega, u \in A^*} \{p(u|\sigma), \mu(\sigma, u) = \sigma\}.$$

Заметим, что $\varrho_S > 0$ для любого неразложимого нетривиального марковского источника S.

В следующей лемме приводится верхняя оценка мощности множества $T_\sigma(q)$.

Лемма 1. *Пусть S — марковский источник в алфавите A с множеством состояний Ω и $\varrho = \varrho_S > 0$. Тогда при любых $\sigma \in \Omega$ и q, $0 < q < 1$, справедливо неравенство*

$$|T_\sigma(q)| \leqslant \frac{4|A|^2|\Omega|^2}{q\varrho}.$$

Доказательство. Рассмотрим дерево

$$T' = M_\sigma(q) \cup \left(\{u \in A^* \mid p(u|\sigma) \geqslant \frac{q\varrho}{|\Omega||A|}\} \cap g(T_\sigma(q))\right).$$

По определению $T' \subset T_\sigma(q)$ и $p(u|\sigma) \geqslant q\varrho/|\Omega||A|$ для любого слова $u \in T'$. Тогда из утверждения 2 (a) имеем

$$1 \geqslant \sum_{u \in g(T')} p(u|\sigma) \geqslant |g(T')| \frac{q\varrho}{|\Omega||A|}. \tag{12}$$

Покажем, что в дереве T' нет путей без разветвлений длиннее $|\Omega| + 1$. Действительно, пусть $w, wx_1, wx_1x_2, \ldots, wx_1 \ldots x_k$, где $k > |\Omega|$, — набор слов из T', соответствующий такому пути. Рассмотрим набор состояний

$\tau_1 = \mu(\sigma, w), \tau_2 = \mu(\sigma, wx_1), \ldots, \tau_k = \mu(\sigma, wx_1 \ldots x_{k-1})$. Найдутся i, j такие, что $1 \leqslant i < j \leqslant k$, $1 \leqslant j-i \leqslant |\Omega|$ и $\tau_i = \tau_j$, причем $wx_1 \ldots x_i \in M_\sigma(q)$, так как $i < j$. Тогда $p(wx_1 \ldots x_{i-1}|\sigma) \geqslant q$ и из определения ϱ следует неравенство $p(x_i \ldots x_{j-1}|\tau_i) \leqslant 1 - \varrho$. Из (4) получаем

$$p(x_i \ldots x_{j-1}|\tau_i) + \sum_{k=i}^{j-1} \sum_{a \neq x_k} p(x_i \ldots x_{k-1}a|\tau_i) = 1.$$

Отсюда следует, что при некоторых $a \in A$ и $t, i \leqslant t \leqslant j-1$, верно неравенство

$$p(x_i \ldots x_{t-1}a|\tau_i) \geqslant \frac{\varrho}{|A|(j-i)} \geqslant \frac{\varrho}{|A||\Omega|}.$$

Тогда из равенства (3) имеем

$$p(wx_1 \ldots x_{t-1}a|\sigma) \geqslant \frac{q\varrho}{|A||\Omega|}$$

и $wx_1 \ldots x_{t-1} \in M_\sigma(q)$. Следовательно, $wx_1 \ldots x_{t-1}a \in T'$. Противоречие, т. е. в дереве T' нет путей без разветвлений длиннее $|\Omega| + 1$. Тогда из утверждения 2 (d) следует, что

$$|T'|/(|\Omega| + 1) \leqslant 2|g(T')| - 1. \tag{13}$$

Из утверждения 2 (c) и неравенств (12), (13) получаем

$$|T_\sigma(q)| = |A||M_\sigma(q)| + 1 \leqslant |A||T'| + 1 \leqslant 4|A||\Omega||g(T')| \leqslant \frac{4|A|^2|\Omega|^2}{q\varrho}.$$

Лемма 1 доказана.

По вероятности слова получим оценку его аддитивной сложности.

Лемма 2. *Пусть S — марковский источник в алфавите A с множеством состояний Ω, $w \in A^*$ и $\sigma \in \Omega$. Пусть числа I и θ удовлетворяют неравенствам $I \geqslant I_S(w|\sigma)$ и $I > \theta = 4\log^2 I|A|^2|\Omega|^3/\varrho_S$. Тогда*

$$l(w) \leq \frac{I}{\log I}\left(1 + \frac{1 + \log\theta}{\log I} + \frac{\log^2\theta}{\log I(\log I - \log\theta)}\right).$$

Доказательство. Пусть $q = \theta/I$, где по условию $q < 1$. Тогда для любого состояния $\sigma \in \Omega$ из леммы 1 следует неравенство

$$|T_\sigma(q)| \leqslant \frac{I}{|\Omega|\log^2 I}.$$

Из утверждения 2 (b) имеем

$$l(\bigcup_{\sigma\in\Omega} T_\sigma(q)) \leqslant \frac{I}{\log^2 I}. \tag{14}$$

Так как для любого $\sigma \in \Omega$ множество $T_\sigma(q)$ является полным деревом, то слово $w \in A^*$ можно представить в виде такой конкатенации $w = w_1 w_2 \dots w_m$, что $w_i \in g(T_{\tau_i}(q))$ при $i < m$ и $w_m \in T_{\tau_m}(q)$, где $\tau_1 = \sigma$, $\tau_i = \mu(\tau_{i-1}, w_{i-1})$. Тогда из (11) следует, что $p(w_i|\tau_i) < q$ при $i < m$. Из равенств (3) и (6) следует, что

$$p(w|\sigma) = p(w_1|\tau_1)p(w_2|\tau_2)\dots p(w_m|\tau_m) < q^{m-1},$$

$$I \geqslant I_S(w|\sigma) \geqslant (m-1)\log\frac{1}{q}. \tag{15}$$

Отсюда и из неравенств (14), (15) получаем неравенство

$$l(w) \leqslant l(\bigcup_{\sigma\in\Omega} T_\sigma(q)) + m - 1 \leqslant \frac{I}{\log\frac{1}{q}} + \frac{I}{\log^2 I}. \tag{16}$$

Используя равенство

$$\frac{1}{\log\frac{1}{q}} = \frac{1}{\log I - \log\theta} = \frac{1}{\log I}\left(1 + \frac{\log\theta}{\log I} + \left(\frac{\log\theta}{\log I}\right)^2 \left(\frac{1}{1 - \log\theta/\log I}\right)\right),$$

из неравенства (16) получаем утверждение леммы 2.

Лемму 2 нетрудно обобщить на набор слов $w_1, \dots, w_m \in A^*$. А именно, если $\sum\limits_{i=1}^{m} I(w_i|\tau_i) \leqslant I$ для некоторого набора состояний $\tau_1, \tau_2, \dots, \tau_m$ и выполнены условия леммы 2, то

$$l(\{w_1, \dots, w_m\}) \leqslant \frac{I}{\log I}\left(1 + \frac{1 + \log\theta}{\log I} + \frac{\log^2\theta}{\log I(\log I - \log\theta)}\right). \tag{17}$$

Пусть x — бесконечная последовательность букв и $x^n = x_1 \dots x_n$ — ее префикс. Тогда из леммы 2 вытекает

Следствие 1. *Пусть S — неразложимый нетривиальный марковский источник в алфавите A с множеством состояний Ω, $\sigma \in \Omega$ и $x \in A^\infty$. Тогда при $n \to \infty$*

$$l(x^n) \leqslant \frac{I(x^n|\sigma)}{\log I(x^n|\sigma)}(1 + o(1)).$$

Доказательство. Если марковский источник S является неразложимым и нетривиальным, то $\varrho_S > 0$. Среди любого набора из $|\Omega|+1$ состояний найдутся два одинаковых. Следовательно, для произвольных $\sigma \in \Omega$ и $u \in A^{|\Omega|+1}$ имеем $p(u|\sigma) \leqslant 1-\varrho_S$. Тогда из (3) получаем неравенство

$$I(x^n|\sigma) \geqslant \lfloor n/(|\Omega|+1)\rfloor \log \frac{1}{1-\varrho_S},$$

т. е. $I(x^n|\sigma) \to \infty$ при $n \to \infty$. Отсюда и из леммы 2 следует справедливость следствия 1.

Для слова $u \in A^*$ введем обозначение $u^s = \underbrace{u\ldots u}_{s}$ — степень слова u. Рассмотрим марковский источник S первого порядка в алфавите A, удовлетворяющий условию:

$$\text{если } \mu_S(b,w) = b \text{ и } p_S(w|b) > 1/2, \text{ то } w = b^s. \tag{18}$$

Справедлива следующая

Лемма 3. *Пусть S — марковский источник первого порядка в алфавите A, удовлетворяющий условию (18), $\xi = \max\limits_{a\in A}\{1, \lceil -\log\log(1/p(a|a))\rceil\}$ и $w_1,\ldots,w_m$ — набор таких слов в алфавите A длины не менее 2, что две последние буквы каждого слова различны. Если величина I удовлетворяет неравенствам $I \geqslant \sum\limits_{i=1}^{m} I_S(w_i)$ и $I > \theta = 8(\log^2 I)(\xi|A|)^5$, то*

$$l(\{w_1,\ldots,w_m\}) \leqslant \frac{I}{\log I}\left(1 + \frac{1+\log\theta}{\log I} + \frac{\log^2\theta}{\log I(\log I - \log\theta)}\right) + \xi|A|.$$

Доказательство. Пусть $E = \{a \in A \mid p(a|a) > 1/2\}$. Для каждой буквы $a \in E$ определим число $i_a = \lceil -\log\log(1/p(a|a))\rceil$ и введем новые буквы $a(0), a(1), \ldots, a(i_a)$. Буква $a(i)$ в указанном ниже смысле соответствует слову a^{2^i} в алфавите A. Рассмотрим алфавит $A' = A \cup (\cup_{a\in E}\{a(0), a(1), \ldots, a(i_a)\})$. Определим марковский источник S' первого порядка в алфавите A'. Сначала определим вспомогательные величины $\widehat{p}(x|y)$ для всех пар букв $x, y \in A'$.

Если $a \in A\backslash E$, то

$$\widehat{p}(x|a) = \begin{cases} p(x|a) & \text{при } x \in A; \\ 0 & \text{при } x \in A'\backslash A. \end{cases}$$

Если $a \in E$, то

$$\widehat{p}(x|a) = \begin{cases} (p(a|a))^{2^{i_a}} & \text{при } x = a(i_a); \\ (p(a|a))^{2^j}(1 - (p(a|a))^{2^j}) & \text{при } x = a(j), j < i_a; \\ p(x|a) & \text{при } x \in A \backslash \{a\}; \\ 0 & \text{при } x = a \text{ или } x = b(i), \text{ где } b \neq a. \end{cases}$$

Определим $\widehat{p}(x|a(i))$ при $i < i_a$ равенствами

$$\widehat{p}(x|a(i)) = \begin{cases} 0, & \text{если } x = a(j), \text{ где } j \geqslant i; \\ \widehat{p}(x|a) & \text{в остальных случаях.} \end{cases}$$

При $i = i_a$ положим $\widehat{p}(x|a(i_a)) = \widehat{p}(x|a)$. Нетрудно видеть, что для каждой буквы $a \in A$

$$\sum_{x \in A'} \widehat{p}(x|a) = \sum_{x \in A'} \widehat{p}(x|a(i_a)) = 1$$

и при $i < i_a$

$$\sum_{x \in A'} \widehat{p}(x|a(i)) = 1 - (p(a|a))^{2^i}.$$

Теперь определим вероятности букв в источнике S' равенствами

$$p'(x|y) = p_{S'}(x|y) = \begin{cases} \widehat{p}(x|y)/(1 - (p(a|a))^{2^i}), & \text{если } y = a(i),\ i < i_a, \\ \widehat{p}(x|y) & \text{в противном случае.} \end{cases}$$

Пусть $\bar{A}^* \subset A^*$ — множество таких слов в алфавите A, в каждом из которых две последние буквы различны. Зададим инъективное отображение $F : \bar{A}^* \to (A')^*$ следующим образом. Пусть $a \in E$ и $a \neq b$. Определим $F(a^k b) = (a(i_a))^q (a(i_a - 1))^{q_{i_a - 1}} \dots (a(0))^{q_0} b$, если $0 \leqslant k = q_0 2^0 + q_1 2^1 + \dots + q_{i_a - 1} 2^{i_a - 1} + q 2^{i_a}$, где $q_i \in \{0, 1\}$, а q — некоторое целое число. В случае $a \notin E$ положим $F(a^k b) = a^k b$.

Произвольное слово $u \in \bar{A}^*$ можно представить в виде

$$u = a_{j_1}^{k_1} a_{j_2}^{k_2} \dots a_{j_t}^{k_t} a_{j_{t+1}} = a_{j_1} a_{j_1}^{k_1 - 1} a_{j_2} a_{j_2}^{k_2 - 1} \dots a_{j_t} a_{j_t}^{k_t - 1} a_{j_{t+1}}.$$

Здесь k_i могут равняться 1. Пусть

$$F(u) = a_{j_1} F(a_{j_1}^{k_1 - 1} a_{j_2}) F(a_{j_2}^{k_2 - 1} a_{j_3}) \dots F(a_{j_t}^{k_t - 1} a_{j_{t+1}}).$$

Для источника S' вероятности букв были выбраны как условные вероятности подслов в описанном выше разбиении. Покажем, что для любого слова $u \in \bar{A}^*$ справедливо равенство $p(u) = p'(F(u))$. Действительно, если $a \notin E$, то $p(a^k b|a) = p'(a^k b|a)$. Если $a \in E$, то справедливы равенства

$$\begin{aligned} p'(F(a^k b)|a) &= p'((a(i_a))^q (a(i_a-1))^{q_{i_a-1}} \dots (a(0))^{q_0} b|a) \\ &= p'((a(i_a))^q a(j_1)a(j_2)\dots a(j_r)b|a) = p(a|a)^{q2^{i_a}} p(a|a)^{2^{j_1}}(1-p(a|a)^{2^{j_1}}) \end{aligned}$$
$$\times \left(\prod_{j=2}^{r} \frac{p(a|a)^{2^{j_s}}(1-p(a|a)^{2^{j_s}})}{(1-p(a|a)^{2^{j_{s-1}}})}\right) \frac{p(b|a)}{(1-p(a|a)^{2^{j_r}})} = (p(a|a))^k p(b|a) = p(a^k b|a).$$

Используя (3), имеем

$$\begin{aligned} &p'(F(u)) \\ &= p'(a_{j_1}|a_{j_{t+1}}) p'(F(a_{j_1}^{k_1-1}a_{j_2})|a_{j_1}) p'(F(a_{j_2}^{k_2-1}a_{j_3})|a_{j_2}) \dots p'(F(a_{j_t}^{k_t-1}a_{j_{t+1}})|a_{j_t}) \\ &= p(a_{j_1}|a_{j_{t+1}}) p(a_{j_1}^{k_1-1}a_{j_2}|a_{j_1}) p(a_{j_2}^{k_2-1}a_{j_3}|a_{j_2})) \dots p(a_{j_t}^{k_t-1}a_{j_{t+1}}|a_{j_t}))) = p(u). \end{aligned}$$

По определению (6) для любого слова $u \in \bar{A}^*$ получаем равенство

$$I_{S'}(F(u)) = I_S(u). \tag{19}$$

Покажем, что

$$\varrho_{S'} = 1 - \max_{x\in A', v\in(A')^*} \{p'(v|x) \mid \mu_{S'}(x,v) = x\} \geqslant 1/2. \tag{20}$$

Другими словами, покажем, что если x — последняя буква слова $v \in (A')^*$, то $p'(v|x) \leqslant 1/2$.

Рассмотрим три возможных случая: 1) $x \in E$; 2) $x \in A' \backslash A$; 3) $x \in A \backslash E$. Из определения источника S' следует, что если $a \in E$, то $p'(b|a) \leqslant 1/2$ для любого $b \in A'$. Поэтому если буква $a \in E$ содержится в слове xv, то $p'(v|x) \leqslant 1/2$. Предположим, что $x = a(i) \in A' \backslash A$. Тогда по определению источника S' либо буква $a \in E$ содержится в слове v, либо $p'(v|x) = 0$. Предположим, что $x \in A \backslash E$. Тогда если в слове v содержится буква $a(i) \in A' \backslash A$, то либо между буквами x и $a(i)$ содержится буква $a \in E$, либо $p'(v|x) = 0$. Предположим, что $xv \in (A \backslash E)^*$. Тогда $p'(v|x) = p(v|x)$ и из условия (18) имеем либо $p(v|x) \leqslant 1/2$, либо $v = x^s$. Однако неравенство $p(x^s|x) > 1/2$ противоречит тому, что $x \notin E$. Таким образом, неравенство (20) доказано.

Из определения источника S' следует, что $|A'| \leqslant |A|\xi$. Пользуясь неравенством (17) и условиями леммы для набора слов $w_1, w_2, \dots, w_m$, получаем

$$l(\{F(w_1), \dots, F(w_m)\}) \leqslant \frac{I}{\log I}\left(1 + \frac{1+\log\theta}{\log I} + \frac{\log^2\theta}{\log I(\log I - \log\theta)}\right). \tag{21}$$

Здесь $\max\{\theta, \sum\limits_{i=1}^{m} I_{S'}(F(w_i))\} \leqslant I$ и $\theta = 8(\log^2 I)(\xi|A|)^5 < I$.

Нетрудно видеть, что $l(\{w_1, \ldots, w_m\}) \leqslant l(\{F(w_1), \ldots, F(w_m)\}) + L$, где L — аддитивная сложность множества слов вида a^{2^j}, соответствующих буквам алфавита A'. Из определений величин i_a и ξ следует, что $L \leqslant |A|\xi$. Тогда из формул (19) и (21) получаем утверждение леммы 3.

3. Аддитивная сложность слов с запретами на подслова

Пусть $Z \subset A^*$. Через $B(Z)$ обозначим множество, состоящее из всех слов, в которых не содежатся подслова из Z. Рассмотрим случай, когда множество запретов Z является конечным, а $B(Z)$ — бесконечным. Тогда множество $B(Z)$ является регулярным языком (см., [7, гл. I, § 6]). Введем обозначение $B_n(Z) = B(Z) \cap A^n$.

Из предложения, доказанного в [10, предложение 4.7.6], нетрудно убедиться в справедливости следующего утверждения.

Утверждение 3. *Если Z — конечное множество, то существует*

$$\lim_{n\to\infty} \frac{1}{n} \log |B_n(Z)|.$$

Оценим аддитивную сложность слов с запретами на подслова в случае, когда мощность множества $B(Z)$ растет экспоненциально.

Теорема 1. *Пусть Z — множество запретов и $\lim\limits_{n\to\infty} \frac{1}{n} \log |B_n(Z)| > 0$. Тогда при $n \to \infty$*

$$L(B_n(Z)) = \frac{\log |B_n(Z)|}{\log n}(1 + o(1)).$$

Доказательство. Пусть $f_n = \log |B_n(Z)|$. Из условия теоремы следует, что при $n \to \infty$

$$f_n = \alpha n(1 + o(1)), \tag{22}$$

где $\alpha > 0$. Тогда нижняя оценка величины $L(B_n(Z))$ вытекает из утверждения 1.

Выберем $k = k(n)$ такое, что справедливы неравенства

$$|A|\frac{f_n}{\log^3 f_n} \geqslant 2^{f_k} \geqslant \frac{f_n}{\log^3 f_n}. \tag{23}$$

Введем обозначение: $\widehat{B}_n(Z) = \bigcup\limits_{i=0}^{n} B_i(Z)$. Так как множество $\widehat{B}_n(Z)$ является деревом, то из утверждения 2 следует, что при $k \to \infty$

$$L(\widehat{B}_k(Z)) = |\widehat{B}_k(Z)| - 1 \leqslant k|B_k(Z)|(1 + o(1)). \tag{24}$$

Поскольку любое слово $w \in B_n(Z)$ можно представить в виде конкатенации $w = w_1 w_2 \dots w_t$, где $|w_i| = k$ при $i < t$ и $|w_t| \leqslant k$, имеем

$$l(w) \leqslant L(\widehat{B}_k(Z)) + t - 1. \tag{25}$$

Покажем, что при $n \to \infty$

$$L(\widehat{B}_k(Z)) = o\Big(\frac{f_n}{\log f_n}\Big). \tag{26}$$

Из неравенств (23) следует, что $|B_k(Z)| \leqslant |A| f_n / \log^3 f_n$. Тогда, используя асимптотическое равенство (22), получаем

$$\alpha k(1 + o(1)) = f_k \leqslant \log f_n - 3 \log\log f_n + \log |A|,$$

т. е. $k = O(\log f_n)$ при $n \to \infty$.

Отсюда и из (23) и (24) следует, что

$$L(\widehat{B}_k(Z)) = O\Big(\log f_n \frac{|A| f_n}{\log^3 f_n}\Big),$$

т. е. равенство (26) доказано.

Покажем, что при $n \to \infty$ справедливо асимптотическое равенство

$$t = \lceil n/k \rceil = \frac{f_n}{\log f_n}(1 + o(1)). \tag{27}$$

Из (22) следует, что $f_n / f_k = (1 + o(1)) n/k$. Тогда из неравенства (23) получаем $f_k \geqslant \log f_n - 3 \log\log f_n$ и при $n \to \infty$

$$n/k = (1 + o(1)) f_n / f_k \leqslant \frac{f_n}{\log f_n}(1 + o(1)). \tag{28}$$

Так как из (22) следует, что $\log f_n \sim \log n$, то формулы (25)–(28) дают требуемую верхнюю оценку величины $L(B_n(Z))$. Теорема 1 доказана.

4. Аддитивная сложность высоковероятных слов

Рассмотрим некоторый неразложимый непериодический марковский источник S в алфавите A с множеством состояний Ω. Пусть $\sigma \in \Omega$. Все слова $w \in A^n$ упорядочим в порядке убывания их вероятностей $p(w|\sigma)$ и будем отбирать наиболее вероятные до тех пор, пока впервые суммарная вероятность отобранных слов не превзойдет некоторый заранее заданный уровень α, $0 < \alpha < 1$. Множество отобранных слов обозначим через $M_n(\alpha, S, \sigma)$.

Теорема о марковских цепях, аналогичная приведенному ниже утверждению, доказана в [2, гл.12, § 2, теорема 2]. Доказательство полностью переносится на случай марковских источников, если использовать определение энтропии источника, данное во введении.

Утверждение 4. *Для любого* α, $0 < \alpha < 1$,

$$\lim_{n\to\infty} \frac{\log |M_n(\alpha, S, \sigma)|}{n} = H(S),$$

$$\lim_{n\to\infty} \frac{1}{n}(\max_{w\in M_n(\alpha,S,\sigma)} I_S(w|\sigma)) = H(S).$$

В следующей теореме содержится асимптотика для $L(M_n(\alpha, S, \sigma))$.

Теорема 2. *Пусть* S *— неразложимый непериодический марковский источник и* $\sigma \in \Omega$. *Тогда при любом* α, $0 < \alpha < 1$, *и* $n \to \infty$ *справедливо асимптотическое равенство*

$$L(M_n(\alpha, S, \sigma)) = \frac{nH(S)}{\log n}(1 + o(1)).$$

Доказательство. Как указано ранее, энтропия неразложимого непериодического источника S положительна, т. е. $H(S) > 0$. Введем обозначение: $I_n = \max\limits_{w\in M_n(\alpha,S,\sigma)} I_S(w|\sigma)$. Из утверждения 4 следует, что $I_n/n \to H(S) > 0$. Поэтому $\log I_n/\log n \to 1$ при $n \to \infty$. Отсюда и из леммы 2 при $n \to \infty$ получаем

$$L(M_n(\alpha, S, \sigma)) \leqslant \frac{I_n}{\log I_n}(1 + o(1)) = \frac{nH(S)}{\log n}(1 + o(1)).$$

Введем обозначение: $h_n = \log |M_n(\alpha, S, \sigma)|$. Тогда из утверждения 4 следуют пределы $\lim\limits_{n\to\infty} h_n/n = H(S)$ и $\lim\limits_{n\to\infty} \log h_n/\log n = 1$. Отсюда и из утверждения 1 при $n \to \infty$ получаем нижнюю оценку

$$L(M_n(\alpha, S, \sigma)) \geqslant \frac{h_n}{\log h_n}(1 + o(1)) = \frac{nH(S)}{\log n}(1 + o(1)).$$

Теорема 2 доказана.

5. Множество слов с заданным составом подслов

В этом разделе будем рассматривать круговые слова $w \in A^*$, т. е. слова вида $a_{i_{n-t}} a_{i_{n-t+1}} \dots a_{i_n} a_{i_1} \dots a_{i_s}$ будем считать подсловами слова $a_{i_1} \dots a_{i_n}$. Пусть $\sigma \in A^k$. Обозначим через $r_w(\sigma)$ число вхождений подслова σ в слово w. Заметим, что $\sum\limits_{\sigma\in A^k} r_w(\sigma) = |w|$. Упорядочим (например, лексикографически) множество A^k. Введем обозначение: $\mathbf{r}_w = (r_w(\sigma_1), r_w(\sigma_2), \dots, r_w(\sigma_T))$, где $\sigma_i \in A^k$ и $T = |A|^k$.

Пусть $\mathbf{r} = (r(\sigma_1), r(\sigma_2), \dots, r(\sigma_T))$ — произвольный вектор с неотрицательными вещественными компонентами. Введем обозначение: $|\mathbf{r}| = \sum\limits_{i=1}^{T} r(\sigma_i)$. Определим величины $I(\mathbf{r})$, $I_\tau(\mathbf{r})$, $r(\tau)$ при $\tau \in A^{k-1}$ равенствами

$$r(\tau) = \sum_{a \in A} r(\tau a), \tag{29}$$

$$I_\tau(\mathbf{r}) = \sum_{a \in A} r(\tau a) \log \frac{r(\tau)}{r(\tau a)}, \tag{30}$$

$$I(\mathbf{r}) = \sum_{\tau \in A^{k-1}} I_\tau(\mathbf{r}). \tag{31}$$

Заметим, что непрерывная дифференцируемость функции $I(\mathbf{r})$ в точке $\mathbf{r} > \mathbf{0}$ следует из ее определения. Если $\mathbf{r}$ — целочисленный вектор, определим множество

$$M(\mathbf{r}) = \{w \in A^* \mid r_w(\sigma) = r(\sigma) \text{ для всех } \sigma \in A^k\},$$

состоящее из слов с одинаковым составом подслов длины k.

Для оценки мощности множества $M(\mathbf{r})$ воспользуемся утверждением из [11], доказательство которого имеется также в [1, гл. 17, теорема 4].

Утверждение 5. *Пусть G — связный ориентированный мультиграф, в котором степени захода и исхода каждой вершины одинаковы и для k-й вершины равны r_k, $1 \leqslant k \leqslant n$. Тогда число различных эйлеровых контуров в мультиграфе G равно*

$$\Delta \prod_{k=1}^{n} (r_k - 1)!,$$

где Δ — число прадеревьев (остовных деревьев, ориентированных от корня) с фиксированным корнем в мультиграфе G.

Справедливо следующее

Утверждение 6. *Пусть $M(\mathbf{r}) \neq \varnothing$ и $|\mathbf{r}| = n$. Тогда*

$$|A|^{k-1} \frac{\prod\limits_{\tau \in A^{k-1}} r(\tau)!}{\prod\limits_{\sigma \in A^k} r(\sigma)!} \geqslant |M(\mathbf{r})| \geqslant \frac{1}{(1 + n/|A|^{k-1})^{|A|^{k-1}}} \frac{\prod\limits_{\tau \in A^{k-1}} r(\tau)!}{\prod\limits_{\sigma \in A^k} r(\sigma)!}.$$

Доказательство. Рассмотрим граф де Брейна G_{k-1}. В этом графе каждую дугу, соответствующую слову $\sigma \in A^k$, заменим на $r(\sigma)$ дуг, соединяющих те же вершины. Удалим изолированные вершины. В полученном мультиграфе $G(M)$ дуги пометим теми же буквами, какими они были помечены в ориентированном графе G_{k-1}. Степени захода и исхода вершины мультиграфа $G(M)$, соответствующей слову $\tau \in A^{k-1}$, одинаковы и равны $r(\tau)$, причем число исходящих из нее дуг, помеченных буквой $a \in A$, равно $r(\tau a)$.

Каждому слову $w \in M(\mathbf{r})$ соответствует эйлеров контур в мультиграфе $G(M)$. Зафиксируем произвольную вершину в мультиграфе $G(M)$. Тогда эйлеров контур будет порождать нумерации (в порядке обхода) дуг, выходящих из каждой вершины. Нумерации и соответствующие эйлеровы контуры будем называть эквивалентными, если они совпадают с точностью до перестановки одинаково помеченных дуг. Ясно, что эквивалентным нумерациям соответствуют одинаковые слова. Нетрудно видеть, что число неэквивалентных нумераций равно

$$\prod_{\tau \in A^{k-1}} \frac{r(\tau)!}{\prod\limits_{a \in A} r(\tau a)!} = \frac{\prod\limits_{\tau \in A^{k-1}} r(\tau)!}{\prod\limits_{\sigma \in A^k} r(\sigma)!}.$$

Учитывая, что для определения слова нужно зафиксировать начало контура, из последнего равенства вытекает требуемая верхняя оценка.

В мультиграфе $G(M)$ имеется эйлеров контур, так как $M(\mathbf{r}) \neq \varnothing$. Пусть $V \subseteq A^{k-1}$ — множество вершин в мультиграфе $G(M)$. Из утверждения 5 следует, что число эйлеровых контуров в мультиграфе $G(M)$ не меньше величины $\prod\limits_{\tau \in V} (r(\tau) - 1)!$. Зафиксируем некоторую вершину в качестве начала контура. Каждый эйлеров контур с фиксированным началом входит в класс эквивалентности, состоящий из $\prod\limits_{\sigma \in A^k} r(\sigma)!$ контуров. Поскольку неэквивалентные эйлеровы контуры порождают различные слова (когда начало фиксировано), справедливо неравенство

$$|M(\mathbf{r})| \geqslant \frac{\prod\limits_{\tau \in V} (r(\tau) - 1)!}{\prod\limits_{\sigma \in A^k} r(\sigma)!} = \frac{\prod\limits_{\tau \in A^{k-1}} r(\tau)!}{\prod\limits_{\tau \in V} r(\tau) \prod\limits_{\sigma \in A^k} r(\sigma)!}.$$

Теперь для получения нижней оценки достаточно применить очевидное неравенство

$$\prod_{\tau \in V} r(\tau) \leqslant \left(\frac{n + |A|^{k-1}}{|A|^{k-1}} \right)^{|A|^{k-1}}.$$

Утверждение 6 доказано.

Утверждение 7. *Пусть $M(\mathbf{r}) \neq \varnothing$ и $|\mathbf{r}| = n$. Тогда*

$$I(\mathbf{r}) + (k-1)\log|A| \geqslant \log|M(\mathbf{r})| \geqslant I(\mathbf{r}) - C\log n,$$

где константа $C > 0$ не зависит от n.

Доказательство. Из (29)–(31) следует равенство

$$I(\mathbf{r}) = \sum_{\tau\in A^{k-1}} I_\tau(\mathbf{r}) = \sum_{\tau\in A^{k-1}} \log\left(\frac{r(\tau)^{r(\tau)}}{\prod_{a\in A} r(\tau a)^{r(\tau a)}}\right). \tag{32}$$

Используя формулу Стирлинга $n! = (n/e)^n\sqrt{2\pi n}e^{\theta_n/12n}$, где $0 \leqslant \theta_n \leqslant 1$, получаем неравенство

$$\log\left(\frac{r(\tau)!}{\prod\limits_{a\in A} r(\tau a)!}\right) \geqslant I_\tau(\mathbf{r}) - \Big(\frac{(|A|-1)}{2}\Big)\Big(\log\frac{r(\tau)}{|A|} + \log 2\pi + \frac{1}{6}\Big).$$

Из него и утверждения 6 следует нижняя оценка.

Нетрудно видеть, что $\dfrac{n!}{n_1!\dots n_m!} \leqslant \dfrac{n^n}{n_1^{n_1}\dots n_m^{n_m}}$, где $n = \sum\limits_{i=1}^m n_i$ и $n_i \geqslant 0$ — целые числа ($0^0 = 0! = 1$). Тогда из утверждения 6 и (32) следует, что $\log|M(\mathbf{r})| \leqslant I(\mathbf{r}) + (k-1)\log|A|$. Утверждение 7 доказано.

6. Аддитивная сложность слов с заданным составом подслов

Для получения асимптотики аддитивной сложности слов с фиксированным составом подслов нам понадобится несколько простых утверждений.

С помощью неравенства Йенсена для функции $\log t$ нетрудно доказать

Утверждение 8. *Пусть $p_i, q_i \geqslant 0$ и $\sum\limits_{i=1}^m p_i = \sum\limits_{i=1}^m q_i = 1$. Тогда*

$$(a) \qquad \sum_{i=1}^m p_i \log\frac{1}{p_i} \leqslant \sum_{i=1}^m p_i\log\frac{1}{q_i};$$

$$(b) \qquad \sum_{i=1}^m p_i\log\frac{1}{p_i} \geqslant p_1\log\frac{1}{p_1} + (\sum_{i=2}^m p_i)\log(1/\sum_{i=2}^m p_i).$$

Утверждение 9. *Пусть S — марковский источник $(k-1)$-го порядка и $w \in A^*$. Тогда*

(a) $I_{S,\tau}(w) \geqslant I_\tau(\mathbf{r}_w)$ для любого $\tau \in A^{k-1}$;

(b) $I_S(w) \geqslant I(\mathbf{r}_w)$.

Доказательство. Из (7) и (30) следует, что

$$I_{S,\tau}(w) = -r_w(\tau)\Big(\sum_{a\in A} \frac{r_w(\tau a)}{r_w(\tau)} \log p_S(a|\tau)\Big),$$

$$I_\tau(\mathbf{r}_w) = -r_w(\tau)\Big(\sum_{a\in A} \frac{r_w(\tau a)}{r_w(\tau)} \log \frac{r_w(\tau a)}{r_w(\tau)}\Big).$$

Так как $\sum\limits_{a\in A} p_S(a|\tau) = \sum\limits_{a\in A} r_w(\tau a)/r_w(\tau) = 1$ для любого $\tau \in A^{k-1}$, то из утверждения 8 получаем пункт (a) утверждения 9. Отсюда и из равенств (8), (31) следует пункт (b) утверждения 9. Утверждение 9 доказано.

Утверждение 10. *Пусть* $I_i > 0$ *и* $\sum\limits_{i=1}^{m} I_i = I$. *Тогда при* $I \to \infty$

$$\sum_{i=1}^{m} I_i/\log I_i = \frac{I}{\log I}(1 + o(1)).$$

Доказательство. Без ограничения общности можно считать, что при любом i, $1 \leqslant i \leqslant m'$, справедливо неравенство $I_i \geqslant I/\log^2 I$, а при $2 \leqslant m'+1 \leqslant i \leqslant m$ — неравенство $I_i < I/\log^2 I$. Положим $I_1' = I_1 + \sum\limits_{i=m'}^{m} I_i$. Тогда $I = I_1' + \sum\limits_{i=2}^{m'-1} I_i$ и $I_1'/\log I_1' \geqslant I_1/\log I_1$, поскольку функция $t/\log t$ возрастает при $t \geqslant e$. В этом случае имеем

$$\sum_{i=1}^{m} \frac{I_i}{\log I_i} \leqslant \frac{I_1'}{\log I_1'} + \sum_{i=2}^{m'-1} \frac{I_i}{\log I_i} + (m - m' + 1)\frac{I}{\log^2 I}$$
$$\leqslant \frac{I}{\log I - 2\log\log I} + (m - m' + 1)\frac{I}{\log^2 I}.$$

Отсюда вытекает справедливость утверждения 10.

Как замечено в работе [5], задача об оценке аддитивной сложности набора слов в алфавите, состоящем из одной буквы, сводится к задаче об оценке сложности вычисления набора степеней одной переменной. Таким образом, из [4] следует

Утверждение 11. *Пусть $F = \{a^{s_1}, a^{s_2}, \ldots, a^{s_m}\}$ — набор степеней буквы $a \in A$ и $N = \prod_{i=1}^{m} s_i$. Тогда при $N \to \infty$*

$$l(F) \leqslant \Big(\frac{\log N}{\log\log N} + \log \max_i s_i + m\Big)(1 + o(1)).$$

В случае, когда $k = 1$ и $A = \{0,1\}$, теорема 3 уточняет остаточный член в формуле для величины $L(M(n-k_n, k_n))$, полученной в [5]. Однако новая оценка, по сравнению с оценкой из [5], справедлива в более узкой области значений параметров, а именно, когда $0 < \lim\limits_{n\to\infty} k_n/n < 1$.

Теорема 3. *Пусть A — конечный алфавит, $\mathbf{r}_n$ — последовательность целочисленных векторов размерности $|A|^k$ такая, что $|\mathbf{r}_n| = n$, $M(\mathbf{r}_n) \neq \varnothing$ и*

$$|\mathbf{r}_n/n - \mathbf{r}_0| = o\Big(\frac{\log\log n}{\log n}\Big)$$

для некоторого положительного вектора $\mathbf{r}_0$. Тогда при $n \to \infty$

$$L(M(\mathbf{r}_n)) = \frac{nI(\mathbf{r}_0)}{\log n}\Big(1 + \frac{\beta_n \log\log n}{\log n}\Big),$$

где $1 \leqslant \liminf\limits_{n\to\infty} \beta_n \leqslant \limsup\limits_{n\to\infty} \beta_n \leqslant 2$.

Доказательство. Поскольку функция $I(\mathbf{r})$ дифференцируема в точке $\mathbf{r}_0 > \mathbf{0}$, при $n \to \infty$ имеем

$$|I(\mathbf{r}_n/n) - I(\mathbf{r}_0)| = O(|\mathbf{r}_n/n - \mathbf{r}_0|) = o\Big(\frac{\log\log n}{\log n}\Big).$$

Заметим, что $nI(\mathbf{r}_n/n) = I(\mathbf{r}_n)$. Тогда из утверждения 7 получаем

$$\log|M(\mathbf{r}_n)| = I(\mathbf{r}_n) + O(\log n) = nI(\mathbf{r}_n/n) + O(\log n)$$
$$= n\Big(I(\mathbf{r}_0) + o\Big(\frac{\log\log n}{\log n}\Big)\Big).$$

Так как $\mathbf{r}_0 > \mathbf{0}$, то $I(\mathbf{r}_0) > 0$. Поэтому из утверждения 1 при $n \to \infty$ получаем $L(M(\mathbf{r}_n)) \geqslant \dfrac{nI(\mathbf{r}_0)}{\log n}\Big(1 + (1 + o(1))\dfrac{\log\log n}{\log n}\Big)$.

Рассмотрим марковский источник S $(k-1)$-го порядка в алфавите A и вероятностями букв $p_S(a|\tau) = r_0(\tau a)/r_0(\tau)$ в состояниях $\tau \in A^{k-1}$. Так как по условию теоремы $p_S(a|\tau) > 0$ для всех $a \in A$ и $\tau \in A^{k-1}$, то $\varrho_S > 0$.

Для произвольного слова $w \in M(\mathbf{r}_n)$ имеем

$$I_S(w) = \sum_{\tau\in A^{k-1}} \sum_{a\in A} r_n(\tau a) \log \frac{r_0(\tau)}{r_0(\tau a)}$$

$$= n\Big(\sum_{\tau\in A^{k-1}} \frac{r_n(\tau)}{n} \Big(\sum_{a\in A} \frac{r_n(\tau a)}{r_n(\tau)} \log \frac{r_0(\tau)}{r_0(\tau a)} \Big)\Big) = n\Big(I(\mathbf{r}_0) + o\Big(\frac{\log\log n}{\log n}\Big)\Big).$$

Поэтому из леммы 2 при $I = I_S(w)$ и $n \to \infty$ следует, что

$$l(w) \leqslant \frac{nI(\mathbf{r}_0)}{\log n}\Big(1 + (1+o(1))\frac{2\log\log n}{\log n}\Big).$$

Отсюда и из определения величины $L(M(\mathbf{r}_n))$ получаем верхнюю оценку. Теорема 3 доказана.

Заметим (см., например, [3, раздел 3.6], что величина $I(\mathbf{r_0})$ равняется энтропии источника S, использованного при доказательстве теоремы 3.

В следующей теореме подобная асимптотика приводится при более слабых условиях на последовательность $\mathbf{r}_n$.

Теорема 4. *Пусть A — конечный алфавит, $\mathbf{r}_n$ — последовательность целочисленных векторов размерности $|A|^k$ такая, что $|\mathbf{r}_n| = n$, $M(\mathbf{r}_n) \neq \varnothing$ и $\lim\limits_{n\to\infty} (\log I(\mathbf{r}_n) \log n)/I(\mathbf{r}_n) = 0$. Тогда при $n \to \infty$*

$$L(M(\mathbf{r}_n)) \sim \frac{I(\mathbf{r}_n)}{\log I(\mathbf{r}_n)} \sim \frac{\log|M(\mathbf{r}_n)|}{\log\log|M(\mathbf{r}_n)|}.$$

Доказательство. 1. Для множества $M(\mathbf{r}_n)$ определим марковский источник S_1 $(k-1)$-го порядка в алфавите A с вероятностями букв $p_{S_1}(a|\tau) = p_1(a|\tau) = r_n(\tau a)/r_n(\tau)$ в состояниях $\tau \in A^{k-1}$.

Пусть $w \in M(\mathbf{r}_n)$. Тогда из определения источника S_1 и равенств (6), (30), (31) вытекают равенства

$$I_{S_1}(w|\tau_0) = I_{S_1}(w) = I(\mathbf{r}_n), \tag{33}$$

где τ_0 — суффикс длины $k-1$ слова w.

2. Объединив буквы алфавита A в блоки фиксированной длины, получим новый алфавит B. Вычислив вероятности блоков, построим такой марковский источник S_2 первого порядка в алфавите B, что вероятности слов в алфавите B и соответствующих им слов в алфавите A будут совпадать.

Вначале определим длину блоков. Для каждого состояния $\tau \in A^{k-1}$ рассмотрим множество $A_\tau \subset A^*$, состоящее из слов w таких, что

$\mu_1(\tau, w) = \tau$ (т. е. τ — суффикс слова τw) и справедливо неравенство $p_1(w|\tau) > 1/2$. Обозначим через w_τ кратчайшее слово из множества A_τ, если $A_\tau \neq \emptyset$. Покажем, что $A_\tau \subset \{w_\tau^s \mid s\text{— целое}\}$. Заметим, что элементы множества A_τ являются префиксами друг друга. Пусть $w_1, w_2 \in A_\tau$, $|w_1| \leqslant |w_2|$ и слово w_2' — префикс слова w_2 такой, что $|w_2'| = |w_1|$. Пусть $w_2' \neq w_1$. Тогда из равенства (3) получаем $p_1(w_2'|\tau) \geqslant p_1(w_2|\tau) > 1/2$ и $p_1(w_1|\tau) > 1/2$, что противоречит равенству (4). Пусть $w \in A_\tau$, тогда $w = w_\tau w'$. Так как слово τ является суффиксом слов τw_τ и $\tau w_\tau w'$, то слово τ является также суффиксом слова $\tau w'$, причем из (3) следует, что

$$p_1(w|\tau) = p_1(w_\tau|\tau)p_1(w'|\tau) > 1/2.$$

Поэтому $w' \in A_\tau$ и w_τ — префикс слова w'. Рассуждая по индукции, заключаем, что $w = w_\tau^s$ при некотором целом числе s.

Покажем, что

$$|w_\tau| \leqslant |A|^{k-1} + k - 2. \tag{34}$$

Пусть в слове w_τ найдутся два вхождения некоторого подслова $\delta \in A^{k-1}$ (возможно перекрывающиеся), например, $w_\tau = w_1\delta w_2\delta w_3$. Рассмотрим слово $w_1\delta w_3$. Поскольку $|\delta w_3| \geqslant k-1$, слово τ является суффиксом слова $w_1\delta w_3$ и из равенства (3) получаем

$$\begin{aligned} p_1(w_1\delta w_3|\tau) = p_1(w_1\delta|\tau)p_1(w_3|\delta) \geqslant p_1(w_1\delta|\tau)p_1(w_2\delta|\delta)p_1(w_3|\delta) \\ = p_1(w_\tau|\tau) > 1/2. \end{aligned}$$

Это противоречит определению слова w_τ, т. е. все подслова длины $k-1$ в слове w_τ разные и справедливо неравенство (34). Определим число Γ равенством

$$\Gamma = (k-1)\prod_{\tau \in A^{k-1}} |w_\tau| \leqslant (k-1)(|A|^{k-1} + k - 2)^{|A|^{k-1}} = C, \tag{35}$$

где C — константа, не зависящая от n.

Пусть $B = A^\Gamma$ и $u \in B^*$. Тогда через $u[\Gamma]$ обозначим слово в алфавите A, соответствующее слову u ($|u[\Gamma]| = \Gamma|u|$).

Рассмотрим марковский источник S_2 первого порядка в алфавите B с вероятностями

$$p_{S_2}(a|b) = p_2(a|b) = p_1(a[\Gamma]|b'), \tag{36}$$

где b' — суффикс длины $k-1$ слова $b[\Gamma]$.

3. Покажем, что источник S_2 удовлетворяет условию (18). Пусть $\mu_{S_2}(b, u) = b$, т. е. b — последняя буква в слове u и $p_2(u|b) > 1/2$. Докажем, что слово u есть степень буквы b. Из равенства (36) получаем,

что $p_1(u[\Gamma]|b') = p_2(u|b) > 1/2$, где b' — суффикс длины $k-1$ слова $b[\Gamma]$ и, следовательно, слова $u[\Gamma]$. Тогда $u[\Gamma] \in A_{b'}$ и по доказанному в пункте 2 $u[\Gamma] = w_{b'}^t$ для некоторого целого числа t. Так как число $|w_{b'}|$ делит Γ, то $b[\Gamma] = w_{b'}^{t'}$ и $u[\Gamma] = (b[\Gamma])^{t/t'}$, т. е. $u = b^s$, где $s = t/t'$.

4. Слово w из $M(\mathbf{r}_n)$ представим в виде $w = v[\Gamma]w'$, где $|w'| < \Gamma$ и $v \in B^*$. Из пункта (b) утверждения 2 вытекает, что $l(\bigcup\limits_{i=0}^{\Gamma} A^i) = |\bigcup\limits_{i=0}^{\Gamma} A^i| - 1$. Поэтому

$$l(w) \leqslant l(v) + l(\bigcup_{i=0}^{\Gamma} A^i) + 1 \leqslant l(v) + (|A|^{\Gamma+1} - 1)/(|A| - 1). \tag{37}$$

Из равенства (36) следует, что

$$I_2(v|b_0) = I_1(v[\Gamma]|\tau_0) \leqslant I_1(w|\tau_0), \tag{38}$$

где буква $b_0 \in B$ такова, что τ_0 — суффикс длины $k-1$ слова $b_0[\Gamma]$. Здесь и далее $I_1 = I_{S_1}$ и $I_2 = I_{S_2}$.

5. Рассмотрим множество

$$E = \{d \in B \mid\ -\log\log\frac{1}{p_2(d|d)} \geqslant \log^2 I(\mathbf{r}_n)\},$$

состоящее из букв, образующих длинные серии. Слово $v \in B^*$ разобьем на подслова, выделяя подслова, состоящие из повторений какой-либо буквы из множества E. Число $I_2(v|b_0)$ представим в виде

$$I_2(v|b_0) = I'(v|b_0) + \sum_{d\in E} I_{2,d}(v|b_0), \tag{39}$$

где $I'(v|b_0) = \sum\limits_{c\in B\setminus E} I_{2,c}(v|b_0)$.

Если $d \neq B$, то рассмотрим марковский источник S_3 первого порядка в алфавите $C = (B\setminus E) \cup \{\lambda\}$, где $\lambda \notin B$ играет роль "фиктивной" буквы. Вероятности букв определим равенствами: $p_{S_3}(c|b) = p_3(c|b) = p_2(c|b)$, если $c, b \in C\setminus\{\lambda\}$; $p_{S_3}(c|\lambda) = p_3(c|\lambda) = 1/|C|$ для всех $c \in C$; $p_{S_3}(\lambda|b) = p_3(\lambda|b) = \sum\limits_{d\in E} p_2(d|b)$. Ясно, что для любых слов $v \in (C\setminus\{\lambda\})^*$ и $c \in C\setminus\{\lambda\}$ имеет место равенство $p_3(v|c) = p_2(v|c)$. По определению источника S_3 для единственной добавленной буквы λ и всех $c \in C$ имеем $p_3(c|\lambda) \leqslant 1/2$. Поэтому источник S_3 так же, как и источник S_2, удовлетворяет условию (18). Если $E = B$, то положим $S_3 = S_2$.

Слово $v \in B^*$ представим в виде конкатенации $v = v_1 v_2 \dots v_m$, где либо $v_i \in C^*$, либо $v_i = d^{s_i}$ для некоторой буквы $d \in E$. Набор слов v_i разобьем на несколько подмножеств. Пусть $F_0 = \{v_i \mid v_i \in C^*\}$ и $F_d = \{v_i \mid v_i = d^{s_i}\}$ для каждого $d \in E$. Через f_0 и f_d обозначим число включений слов из множеств F_0 и F_d соответственно в конкатенацию $v_1 v_2 \dots v_m$.

Из введенных обозначений непосредственно вытекают соотношения:

$$f_0 \leqslant \sum_{d\in E} f_d + 1, \quad |F_d| \leqslant f_d \tag{40}$$

и

$$l(v) \leqslant l(F_0) + \sum_{d\in E} l(F_d) + 2\sum_{d\in E} f_d. \tag{41}$$

6. Оценим сверху величину $l(F_d)$ для произвольного $d \in E$. Для упрощения вычислений будем полагать, что d не является последней буквой слова v и каждое слово $v_i \in F_d$ входит в конкатенацию $v_1 v_2 \dots v_m$ однократно, т. е. $f_d = |F_d|$. Пусть $N_d = \prod_{v_i\in F_d} |v_i|$. Из утверждения 11 следует, что при $N_d \to \infty$

$$l(F_d) \leqslant (1+o(1))\Big(\frac{\log N_d}{\log\log N_d} + \log\frac{n}{\Gamma} + |F_d|\Big). \tag{42}$$

Покажем, что

$$\log N_d \leqslant I_{2,d}(v|b_0). \tag{43}$$

Учитывая неравенство $\sum\limits_{v_i\in F_d} |v_i| \leqslant r_v(d)$, как и в [5] получаем

$$\log\prod_{v_i\in F_d} |v_i| \leqslant \log\Big(\frac{r_v(d)}{f_d}\Big)^{f_d} = f_d\log\frac{r_v(d)}{f_d}.$$

Так как $f_d = \sum_{b\neq d} r_v(db)$, то из утверждений 8 ($b$) и 9 ($a$) следует, что

$$f_d\log\frac{r_v(d)}{f_d} \leqslant I_d(\mathbf{r}_v) \leqslant I_{2,d}(v) \leqslant I_{2,d}(v|b_0).$$

Теперь покажем от противного, что при $n \to \infty$ справедливо неравенство

$$l(F_d) \leqslant (1+o(1))\Big(\frac{I_{2,d}(v|b_0)}{\log I_{2,d}(v|b_0)} + \log n + f_d\Big) + \log I(\mathbf{r}_n). \tag{44}$$

Пусть при $n \to \infty$ неравенство (44) не выполняется. Тогда найдется подпоследовательность n_m такая, что $l(F_d) \geqslant \log I(\mathbf{r}_{n_m})$, причем по условию теоремы $\log I(\mathbf{r}_{n_m}) \to \infty$ при $m \to \infty$. Следовательно, $l(F_d) \to \infty$ и

$N_d \to \infty$ при $m \to \infty$, так как $\max_{v_i \in F_d} \log |v_i| \geqslant l(F_d)$. Тогда неравенство (44) получается из неравенств (42), (43) и возрастания функции $\frac{t}{\log t}$ при $t \geqslant e$. Противоречие. Таким образом справедливость неравенства (44) установлена.

7. Оценим сверху величину f_d для произвольной буквы $d \in E$. Из неравенства $t' - 1 \geqslant \ln t'$ при $t' > 0$ нетрудно получить справедливое при $t \in (0,1)$ неравенство $\ln \frac{1}{1-t} \geqslant -\ln \ln \frac{1}{t}$ и неравенство $\log \frac{1}{1-t} \geqslant -\log \log \frac{1}{t} + \log \log e$. Из пункта ($b$) утверждения 8 получаем

$$I_{2,d}(v|b_0) = \sum_{b \in B} r_v(db) \log \frac{1}{p_2(b|d)} \geqslant f_d \log \frac{1}{1 - p_2(d|d)} \geqslant f_d \Big(-\log \log \frac{1}{p_2(d|d)} + \log \log e \Big).$$

Тогда из определения множества E и формул (33), (38), (39) следуют неравенства

$$I_{2,d}(v|b_0) \leqslant I(\mathbf{r}_n) \tag{45}$$

и

$$f_d \leqslant \frac{I_{2,d}(v|b_0)}{-\log \log (p_2(d|d))^{-1} + \log \log e} \leqslant \frac{I(\mathbf{r}_n)}{\log^2 I(\mathbf{r}_n)}. \tag{46}$$

8. Используя лемму 3, оценим сверху величину $l(F_0)$. Обозначим через v_i' слово v_i без первой буквы, т. е. $v_i = c_i v_i'$. Из равенства (39) следует, что

$$I'(v|b_0) = -\sum_{v_i \in F_0} \log p_2(v_i' d_i | c_i),$$

где d_i — буква, следующая за подсловом v_i в слове v. Из определения множества F_0 следует, что $d_i \notin C$, а из определения источника S_3 следует, что $-\log p_3(\lambda|c) \leqslant -\log p_2(d_i|c)$ для произвольной буквы $c \in C$. Поэтому

$$I'(v|b_0) \geqslant -\sum_{v_i \in F_0} \log p_3(v_i' \lambda | c_i) = -\sum_{v_i \in F_0} \log p_3(v_i \lambda | \lambda) + \log p_3(c_i|\lambda).$$

Отсюда и из определения источника S_3 следует, что

$$\sum_{v_i \in F_0} I_3(v_i \lambda) = \sum_{v_i \in F_0} I_3(v_i \lambda | \lambda) \leqslant I'(v|b_0) + f_0 \log \frac{1}{|C|}.$$

Так как буква λ отсутствует в словах v_i, то всюду сократив ее в схеме конкатенации слов $v_i \lambda$, получим схему конкатенации слов v_i. Следовательно,

$l(F_0) \leqslant l(\{v_i\lambda | v_i \in F_0\})$. Пусть $\widehat{I} = I'(v|b_0) + f_0 \log \frac{1}{|C|}$, $h_n = \log I(\mathbf{r}_n)$ и $I = \max\{\widehat{I}, h_n^{1+\log\log h_n}\}$. При $n \to \infty$ имеем

$$8(\log^2 I(\mathbf{r}_n)|B|)^5 < \frac{h_n^{1+\log\log h_n}}{((1+\log\log h_n)\log h_n)^2} \leqslant \frac{I}{\log^2 I},$$

так как по условию теоремы $h_n \to \infty$. Таким образом, выполнены условия леммы 3, из которой следует неравенство

$$l(F_0) \leqslant \frac{I}{\log I}\Big(1 + \frac{1+\log\theta}{\log I} + \frac{\log^2\theta}{\log I(\log I - \log\theta)}\Big) + |B|\log^2 I(\mathbf{r}_n),$$

где $\theta = 8\log^2 I(\log^2 I(\mathbf{r}_n)|B|)^5$. Так как

$$\log\log I(\mathbf{r}_n) = o(\log h_n(1+\log\log h_n)) = o(1)\log I,$$

то $\log\theta = o(1)\log I$ и предыдущее неравенство можно представить в виде

$$l(F_0) \leqslant \frac{I}{\log I}(1+o(1)) + |B|\log^2 I(\mathbf{r}_n). \tag{47}$$

Из определения величины h_n следует, что при $n \to \infty$

$$h_n^{1+\log\log h_n} = o\Big(\frac{I(\mathbf{r}_n)}{\log I(\mathbf{r}_n)}\Big).$$

Тогда, учитывая определения величин I и $\widehat{I}$, из неравенств (40), (46) и (47) при $n \to \infty$ получаем

$$l(F_0) \leqslant \left(\frac{\widehat{I}}{\log \widehat{I}} + \frac{h_n^{1+\log\log h_n}}{\log h_n(1+\log\log h_n)}\right)(1+o(1)) =$$

$$\frac{I'(v|b_0)}{\log I'(v|b_0)} + o\Big(\frac{I(\mathbf{r}_n)}{\log I(\mathbf{r}_n)}\Big). \tag{48}$$

9. Получим верхнюю оценку величины $l(w)$ для слов $w \in M(\mathbf{r}_n)$. Пусть слово $v \in B^*$ соответствует слову $w \in A^*$ как указано в п. 4. Из неравенств (41), (44), (45), (46) и (48) следует, что при $n \to \infty$

$$l(v) \leqslant \frac{I'(v|b_0)}{\log I'(v|b_0)} + \sum_{d\in E}\frac{I_{2,d}(v|b_0)}{\log I_{2,d}(v|b_0)} + o\Big(\frac{I(\mathbf{r}_n)}{\log I(\mathbf{r}_n)}\Big) + C\log n,$$

где $C > 0$ не зависит от n. Из условия теоремы следует, что при $n \to \infty$

$$\log n = o\Big(\frac{I(\mathbf{r}_n)}{\log I(\mathbf{r}_n)}\Big). \tag{49}$$

Используя утверждение 10 и формулу (39), при $n \to \infty$ получаем

$$l(v) \leqslant \frac{I_2(v|b_0)}{\log I_2(v|b_0)}(1 + o(1)) + o\Big(\frac{I(\mathbf{r}_n)}{\log I(\mathbf{r}_n)}\Big).$$

Отсюда и из (33), (37), (38) следует, что при $n \to \infty$

$$l(w) \leqslant \frac{I(\mathbf{r}_n)}{\log I(\mathbf{r}_n)}(1 + o(1)). \tag{50}$$

10. Из (49) и утверждения 7 следует соотношение $\log |M(\mathbf{r}_n)| \sim I(\mathbf{r}_n)$. Отсюда и из утверждения 1 при $n \to \infty$ получаем

$$L(M(\mathbf{r}_n)) \geqslant \frac{I(\mathbf{r}_n)}{\log I(\mathbf{r}_n)}(1 + o(1)).$$

Пользуясь этим неравенством и (50), при $n \to \infty$ имеем

$$L(M(\mathbf{r}_n)) \sim \frac{I(\mathbf{r}_n)}{\log I(\mathbf{r}_n)}.$$

Теорема 4 доказана.

Пусть x — последовательность букв в алфавите A и слово $x^n = x_1 \dots x_n$ — ее префикс длины n. Рассмотрим вектор $\mathbf{r}_n(k, x) = (r_{x^n}(\sigma_1), \dots, r_{x^n}(\sigma_T))$, где $\sigma_i \in A^k$ и $T = |A|^k$. Тогда $x^n \in M(\mathbf{r}_n(k, x))$ и $l(x^n) \leqslant L(M(\mathbf{r}_n(k, x)))$. Из теоремы 4 непосредственно вытекает

Следствие 2. *Пусть* $\lim\limits_{n\to\infty} (\log I(\mathbf{r}_n(k, x)) \log n)/I(\mathbf{r}_n(k, x)) = 0$ *при* $x \in A^\infty$. *Тогда при* $n \to \infty$

$$l(x^n) \leqslant \frac{I(\mathbf{r}_n(k, x))}{\log I(\mathbf{r}_n(k, x))}(1 + o(1)).$$

ЛИТЕРАТУРА

1. **Берж К.** Теория графов и ее применения. М.: Изд-во иностранной литературы, 1962.
2. **Боровков А. А.** Теория вероятностей. М.: Наука, 1976.

3. **Галлагер Р.** Теория информации и надежная связь. М.: Советское радио, 1974.

4. **Гашков С. Б., Кочергин В. В.** Об аддитивных цепочках векторов, вентильных схемах и сложности вычисления степеней // Методы дискретного анализа в теории графов и сложности: Сб. науч. тр. Вып. 52. Новосибирск: Ин-т математики СО РАН, 1992. С. 24–40.

5. **Кочергин В. В.** О мультипликативной сложности двоичных слов с заданным числом единиц // Математические вопросы кибернетики. Вып. 8. М.: Наука, 1999. С. 63–76.

6. **Лупанов О. Б.** Об одном подходе к синтезу управляющих систем — принципе локального кодирования // Проблемы кибернетики. Вып. 14. М.: Наука, 1965. С. 31–110.

7. **Марков А. А.** Введение в теорию кодирования. М.: Наука, 1982.

8. **Потапов В. Н.** Энтропия и мультипликативная сложность слов // Материалы XI Межгосударственной школы-семинара "Синтез и сложность управляющих систем" (Нижний Новгород, 20–25 ноября 2000 г.). Часть II. М.: Изд-во центра прикл. исслед. при мех.-мат. фак. МГУ, 2001. С. 160–162.

9. **Потапов В. Н.** Энтропийная оценка мультипликативной сложности // Проблемы теоретической кибернетики: Тезисы докладов XIII Международной конференции (Казань, 27–31 мая 2002 г.). Часть II. М.: Изд-во центра прикл. исслед. при мех.-мат. фак. МГУ, 2002. С. 154.

10. **Стенли Р.** Перечислительная комбинаторика. М.: Мир, 1990.

11. **van Aardenne-Ehrenfest T., de Bruijn N. G.** Circuits and trees in oriented linear graphs // Simon Stevin. 1951. V. 28. P. 203–217.

Адрес автора:

Институт математики
им. С. Л. Соболева СО РАН,
пр. Академика Коптюга, 4,
630090 Новосибирск,
Россия
E-mail: vpotapov@math.nsc.ru

Статья поступила
15 августа 2003 г.