ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Октябрь—декабрь 1996. Том 3, № 4, 69–76

УДК 519.174

О ЛОКАЛЬНО ИЗОМЕТРИЧЕСКОМ КОДИРОВАНИИ НАТУРАЛЬНЫХ ЧИСЕЛ*)

А. Л. Пережогин

Обратимое отображение $f : N_l \longrightarrow \{0,1\}^n$ множества первых $l+1$ натуральных чисел $N_l = \{0,1,\dots,l\}$ в множество двоичных наборов длины n называется m-изометрическим кодированием, $1 \leqslant m \leqslant n$, если расстояние Хемминга между наборами $f(i)$ и $f(j)$ равно $|i-j|$ для всех $i,j \in N_l$ таких, что $|i-j| \leqslant m$. В настоящей работе для произвольного n и $m = n-1$ получена экспоненциальная нижняя оценка наибольшей длины кодируемого отрезка натурального ряда и найдена конструкция кода, для которого эта оценка является точной при некоторых дополнительных ограничениях.

Пусть $N_l = \{0,1,\dots,l\}$. Обозначим через I^n множество наборов длины n в алфавите $I = \{0,1\}$. Обратимое отображение $f : N_l \longrightarrow I^n$ называется *m-изометрическим*, $m \geqslant 1$, если

$$\rho(f(i), f(j)) = |i-j|$$

для всех $i,j \in N_l$ таких, что $|i-j| \leqslant m$, где ρ — метрика Хемминга на множестве I^n.

Следуя работам [1,2], будем рассматривать f как такое отображение, которое кодирует числа множества N_l двоичными наборами длины n и обладает свойством локальной изометричности, т. е. сохраняет те расстояния, которые не превосходят m. При исследовании изометрических кодирований возникает следующая задача.

$(*)$ Пусть заданы натуральные числа n и m, где $m < n$. Найти максимально возможную длину l отрезка натурального ряда, для которого существует m-изометрическое отображение $f : N_l \longrightarrow I^n$. Какова конструкция последовательности кодовых наборов $f(0),\ f(1),\dots,f(l)$?

Поскольку f обратимо и $m \geqslant 1$, в задаче $(*)$ максимальное значение l не превосходит $2^n - 1$. При $l = 2^n - 1$ взаимно однозначное отображение f^{-1} можно считать нумерацией множества двоичных наборов длины

*) Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 96–01–01800).

n натуральными числами $0, 1, \ldots, 2^n - 1$. Впервые m-изометрические отображения введены и исследованы в работе [3], где и были названы $< m, n >$-нумерациями длины l.

В частности, в этой же работе доказано, что m-изометрические отображения $f : N_l \longrightarrow I^n$ существуют для любых $n, m \leqslant n/2+1$ и $l = 2^n - 1$ за исключением случая $n = 4$, $m = 3$; в этом случае наибольшее значение l равно 13. Ниже мы исследуем задачу $(*)$ для $m = n - 1$ и показываем, что $(n-1)$-изометрическое отображение $f : N_l \longrightarrow I^n$ существует для любого $l \leqslant (n-1)2^{\lceil n/2 \rceil} - 1$, причем эта верхняя оценка точна при некоторых дополнительных ограничениях.

Как и в работе [3], мы будем использовать редукцию задачи $(*)$ к вопросу о существовании и построении символьных последовательностей с ограничениями на их подслова и на вхождения символов в эти подслова. (Подсловом длины k слова X называется слово, состоящее из k последовательных букв в X.)

Обозначим через $L_{n,m}$ множество всех слов в алфавите $A_n = \{1, 2, \ldots, n\}$, удовлетворяющих следующим условиям:

(i) в каждом подслове длины m все буквы различны;

(ii) в каждое подслово произвольной длины некоторая буква входит нечетное число раз.

Для любого множества слов L положим

$$d(L) = \max_{X \in L} |X|, \text{ где } |X| \text{ — длина слова } X.$$

В [3] показано, что задача $(*)$ равносильна задаче о нахождении $d(L_{n,m})$ и конструировании слов из $L_{n,m}$, т. е. слов, удовлетворяющих условиям (i) и (ii).

Введем ряд определений и обозначений.

Если X — слово в алфавите A_n, то набор $S(X) = (s_1(X), s_2(X), \ldots, s_n(X))$ называется *набором четности* для X, где

$$s_i(X) = \begin{cases} 0, & \text{если } i\text{-я буква алфавита } A_n \\ & \text{входит в } X \text{ четное число раз,} \\ 1 & \text{в противном случае.} \end{cases}$$

Заметим, что если $X = YZ$, то $S(X) = S(Y) \oplus S(Z)$, где $\oplus$ — операция покомпонентного сложения по модулю 2.

Слово X называется *циклическим*, если $S(X) = (0, 0, \ldots, 0)$. Циклическое слово будем считать записанным по окружности, т. е. за последней буквой слова X следует его первая буква.

Обозначим через $L^*_{n,m}$ множество циклических слов длины l в алфавите A_n, которые удовлетворяют условию (i) и в каждое подслово произвольной длины, меньшей l, некоторая буква входит нечетное число раз.

Нетрудно видеть, что если для циклического слова X выполнено условие (i), то последнее свойство равносильно тому, что в X существует подслово длины $l-1$, удовлетворяющее условию (ii).

Из определений множеств $L_{n,m}$ и $L^*_{n,m}$ следует, что

$$d(L_{n,m}) \geqslant d(L^*_{n,m}) - 1.$$

Для слов $X = x_1\, x_2\, \ldots\, x_l$ и $Y = y_1\, y_2\, \ldots\, y_l$ введем обозначение

$$\langle X, Y\rangle = x_1\; y_1\; x_2\; y_2\; \ldots\; x_l\; y_l.$$

Для произвольных $X = x_1\, x_2\, \ldots\, x_l \in L^*_{n,m}$ и $j \in A$ пусть

$$IN_j(X) := \{i \in N_l' \;: x_i = j\},$$

где $N_l' = N_l \backslash \{0\}$.

Для слова $X = x_1\, x_2\, \ldots\, x_l$ из $L^*_{n,m}$ определим следующее отображение $prev_X : N_l' \longrightarrow N_l'$.

Пусть $p \in N_l'$ и $IN_{x_p}(X) = \{i_1, i_2, \ldots, i_{m_j}\}$, где $i_1 < i_2 < i_3 < \ldots < i_{m_j}$. Тогда существует $b \in \{1, 2, \ldots, m_j\}$ такое, что $p = i_b$. Положим

$$prev_X(p) = \begin{cases} i_{b-1}, & \text{если } b \geqslant 2, \\ i_{m_j}, & \text{если } b = 1. \end{cases}$$

Таким образом, функция $prev_X$ отображает номер каждого вхождения произвольной буквы слова X в номер предыдущего по циклу вхождения этой же буквы. Например, если $X = x_1\, x_2\, \ldots\, x_8 = 1\,2\,1\,3\,1\,2\,1\,3$, то

$$prev_X(1) = 7,\; prev_X(2) = 6,\; prev_X(3) = 1,\; prev_X(4) = 8,$$
$$prev_X(5) = 3,\; prev_X(6) = 2,\; prev_X(7) = 5,\; prev_X(8) = 4.$$

Пусть для любого $i \in N_l'$

$$\mu_X(i) = \begin{cases} 1, & \text{если число } prev_X(i) - i \text{ является четным}, \\ 0 & \text{в противном случае}. \end{cases}$$

Обозначим $\nu(X) = \sum\limits_{i=1}^{l} \mu_X(i)$ и $\delta(X) = \nu(X)/|X|$. Легко видеть, что $0 \leqslant \delta(X) \leqslant 1$.

Для приведенного выше примера имеем $\delta(1\,2\,1\,3\,1\,2\,1\,3) = 1$.

Рассмотрим множество

$$Q_n = \{X \in L^*_{n,n-1} \mid \delta(X) = 1\}.$$

Теорема 1. *При любом нечетном $n \geqslant 3$ справедливо неравенство*

$$d(Q_n) \leqslant (n-1)2^{(n+1)/2}.$$

Доказательство. Пусть $X = x_1 x_2 \dots x_l$ — произвольное слово из Q_n, $n \geqslant 3$ и нечетно. Полагая $X = \langle Y, Z\rangle$, рассмотрим слова $Y = y_1 y_2 \dots y_{l/2}$ и $Z = z_1 z_2 \dots z_{l/2}$, где

$$y_i = x_{2i-1}, \quad z_i = x_{2i}, \qquad 1 \leqslant i \leqslant l/2.$$

Поскольку $\delta(X) = 1$, то $y_i \neq z_j$ для любых $i, j \in \{1, 2, \dots, l/2\}$, т. е. каждая буква из A_n встречается лишь в одном из слов Y или Z. Так как $X \in L^*_{n,n-1}$, то в каждом слове Y и Z любые $(n-1)/2$ последовательные буквы различны. Следовательно, одно из слов (пусть Y) является многократным повторением блока длины $(n-1)/2$ различных букв из A_n, т. е. имеет вид

$$Y = a_1 a_2 \dots a_{(n-1)/2}\, a_1 a_2 \dots a_{(n-1)/2} \dots a_1 a_2 \dots a_{(n-1)/2}, \tag{1}$$

где $a_i \in A_n$. Рассмотрим отображение φ такое, что для любого набора $V = (v_1, v_2, \dots, v_n)$

$$\varphi(V) = (v_{a_1}, v_{a_2}, \dots, v_{a_{(n-1)/2}}).$$

Так как $X \in L^*_{n,n-1}$, то в упорядоченном списке двоичных наборов $\Theta = \{W_1, W_2, \dots, W_l\} = \{S(x_1), S(x_1 x_2), \dots, S(x_1 x_2 \dots x_l)\}$ все наборы различны. Рассмотрим упорядоченный список наборов $\varphi(\Theta) = \{\varphi(W_1), \varphi(W_2), \dots, \varphi(W_l)\}$. Так как слово Z не содержит букв из множества $\{a_1, a_2, \dots, a_{(n-1)/2}\}$, то

$$\varphi(W_{2i}) = \varphi(W_{2i-1}) \tag{2}$$

при всех $i \in \{1, 2, \dots, l/2\}$.

Так как слово Y имеет вид (1), в любом подслове длины $2(n-1)$ слова X каждая буква из $\{a_1, a_2, \dots, a_{(n-1)/2}\}$ встречается дважды. Поэтому

$$\begin{gathered}\varphi(W_i) = \varphi(W_{i-(2n-2)}) \text{ при всех } i \in N'_l \backslash N'_{2n-3},\\ \varphi(W_i) = \varphi(W_{i-(2n-2)+l}) \text{ при всех } i \in N'_{2n-3}.\end{gathered}$$

Из (2) и (3) следует, что в списке $\varphi(\Theta)$ имеется не более $n-1$ различных наборов. Но в списке Θ все наборы различны, а для любого набора $U \in I^{(n-1)/2}$ имеется только $2^{(n+1)/2}$ различных наборов U' таких, что $\varphi(U') = U$. Значит, $l \leqslant (n-1)2^{(n+1)/2}$. Теорема 1 доказана.

Теорема 2. *При любом нечетном $n \geqslant 3$ справедливо неравенство*

$$d(L^*_{n,n-1}) \geqslant (n-1)2^{(n+1)/2}.$$

Доказательство. Воспользуемся индукцией по нечетным n. Легко проверить, что для $n = 3$ слово $H_3 = 1\,3\,2\,3\,1\,3\,2\,3$ искомое.

Пусть для некоторого нечетного $n \geqslant 3$ построено слово H_n из $L^*_{n,n-1}$ такое, что

$$|H_n| = (n-1)2^{(n+1)/2}. \qquad (4)$$

Кроме того, пусть

$$H_n = U_1^n \; n \; U_2^n \; n \; \ldots \; U_{l(n)}^n \; n \qquad (5)$$

(здесь и в дальнейшем $l(n) = 2^{(n+1)/2}$), где $|U_i^n| = n-2$, $1 \leqslant i \leqslant l(n)$, и пусть

$$U_i^n = U_{i+l(n)/2}^n, \qquad 1 \leqslant i \leqslant l(n)/2. \qquad (6)$$

Заметим, что H_3 удовлетворяет всем этим условиям.

Очевидно, что слово U_m^n является словом в алфавите A_{n-1} и все буквы в U_m^n различны. Значит, существует только одна буква из A_{n-1}, которой нет в слове U_m^n. Обозначим ее через u_m^n.

Рассмотрим следующее слово в алфавите A_{n+2}:

$$H_{n+2} = n \, V_1^n \, n \, V_2^n \ldots n \, V_{l(n)}^n,$$

где при любом $i \in \{1, 2, \ldots, l(n)\}$

$$V_i^n = n+1 \; U_i^n \; n+2 \; u_i^n \; n+1 \; U_i^n \; n+2$$

(здесь и в дальнейшем записи $n+1$, $n+2$ обозначают буквы). Тогда из (6) следует, что

$$V_i^n = V_{i+l(n)/2}^n, \quad 1 \leqslant i \leqslant l(n)/2. \qquad (7)$$

Значит, H_{n+2} является циклическим.

Заметим, что H_{n+2} можно переписать в виде

$$n+2 \; U_1^{n+2} \; n+2 \; U_2^{n+2} \; \ldots \; n+2 \; U_{2l(n)}^{n+2},$$

где

$$U_{2j-1}^{n+2} = n \; n+1 \; U_j^n, \qquad U_{2j}^{n+2} = u_j^n \; n+1 \; U_j^n, \qquad 1 \leqslant j \leqslant l(n).$$

Очевидно, что H_{n+2} удовлетворяет условиям (4) и (i). Поэтому для доказательства включения $H_{n+2} \in L^*_{n+2,n+1}$ достаточно показать, что некоторое подслово длины $(n+1)2^{(n+3)/2} - 1$ слова $H_n + 2$ удовлетворяет условию (ii).

Покажем, что подслово

$$V_1^n \; n \; V_2^n \; n \; \ldots \; n \; V_{l(n)}^n \qquad (8)$$

удовлетворяет условию (ii).

Предположим противное, т. е. в слове (8) существует подслово M такое, что $S(M) = (0, 0, \ldots, 0)$.

В дальнейшем верхний индекс n не пишем.

Подслова слова (8) имеют один из следующих видов:

(I) $V_i\ n\ \dots\ n\ V_j$,

(II) $n\ V_i\ n\ \dots\ n\ V_j\ n$,

(III) $n\ V_i\ n\ \dots\ n\ V_j$,

(IV) $V_i\ n\ \dots\ n\ V_j\ n$,

(V) $Z_1\ n\ V_i\ n\ \dots\ n\ V_j$,

(VI) $Z_1\ n\ V_i\ n\ \dots\ n\ V_j\ n$,

(VII) $V_i\ n\ \dots\ n\ V_j\ n\ Z_2$,

(VIII)$n\ V_i\ n\ \dots\ n\ V_j\ n\ Z_2$,

(IX) $Z_1\ n\ V_i\ n\ \dots\ n\ V_j\ n\ Z_2$,

где Z_1 — непустой суффикс слова V_{i-1}, $|Z_1| < |V_{i-1}|$, Z_2 — непустой префикс слова V_{i+1}, $|Z_2| < |V_{i+1}|$.

Рассмотрим каждый из этих случаев и покажем, что $S(M) \neq (0,0,\dots,0)$. Легко видеть, что для любого $i \in N'_{l(n)}$

$$S(V_i) = S(u_i). \tag{9}$$

В случаях (I) и (II) длина M нечетна, поэтому $S(M) \neq (0,0,\dots,0)$.

Рассмотрим случай (III). Буква n входит в M четное число раз. Значит, число $i-j$ является нечетным. Из (9) имеем

$$(0,0,\dots,0) = S(M) = S(V_i\ \dots\ V_j) = S(u_i\ u_{i+1}\ \dots\ u_j). \tag{10}$$

Так как число $j-i$ является нечетным, то

$$S(U_i\ u_i\ \dots\ U_j\ u_j) = (0,0,\dots,0).$$

Отсюда с использованием (10) получаем

$$\begin{aligned}(0,0,\dots,0) = S(U_i\ u_i\ \dots\ U_j\ u_j) &= S(u_i\ \dots\ u_j) \oplus S(U_i\ \dots\ U_j)\\ &= S(U_i\ \dots\ U_j) = S(n\ U_i\ n\ \dots\ n\ U_j) \neq (0,0,\dots,0).\end{aligned}$$

Аналогично рассматривается случай (IV).

В случаях (V), (VI), (VII) и (VIII) $s_{n+1}(M) = 1$ или $s_{n+2}(M) = 1$.

Рассмотрим случай (IX). Так как $s_n(M) = 0$, то число $j-i$ является четным.

Буквы $n+1$ и $n+2$ входят в слово $n\ V_i\ n\ V_{i+1}\ n\ \dots\ n\ V_j\ n$ четное число раз. Значит, и в слово $Z_1\,Z_2$ они тоже входят четное число раз. Но тогда для Z_1 и Z_2 возможны только следующие случаи:

(a) $Z_1 = X\ n+2\ u_{i-1}\ n+1\ U_{i-1}\ n+2$ и $Z_2 = n+1\ Y$,

(b) $Z_1 = X\ n+2$ и $Z_2 = n+1\ U_{j+1}\ n+2\ u_{j+1}\ n+1\ Y$,

(c) $Z_1 = u_{i-1}\ n+1\ U_{i-1}\ n+2$ и $Z_2 = n+1\ U_{j+1}\ n+2$,

(d) $Z_1 = n+1\ U_{i-1}\ n+2$ и $Z_2 = n+1\ U_{n+1}\ U_{j+1}\ n+2\ u_{j+1}$,

(e) $Z_1 = n+1\ U_{i-1}\ n+2$ и $Z_2 = n+1\ U_{j+1}\ n+2$,

(f) $Z_1 = u_{i-1}\ n+1\ U_{i-1}\ n+2$ и $Z_2 = n+1\ U_{j+1}\ n+2\ u_{j+1}$,

где X — суффикс слова U_{i-1}, $0 \leqslant |X| \leqslant n-2$, Y — префикс слова U_{j+1}, $0 \leqslant |Y| \leqslant n-2$.

Рассмотрим случай (IXa). Используя (9), имеем

$$\begin{aligned}(0,0,\dots,0) = S(M) = S(&Z_1\ u_i\ u_{i+1}\ \dots\ u_j\ Z_2)\\ &= S(X\ u_{i-1}\ U_{i-1}\ u_i\ u_{i+1}\ \dots\ u_j\ Y).\end{aligned} \tag{11}$$

Поскольку число $i-j$ является четным, то

$$S(u_{i-1}\ U_{i-1}\ u_i\ U_i\ \dots\ u_j\ U_j) = (0,0,\dots,0).$$

Тогда, используя (11), имеем

$$\begin{aligned}(0,0,...,0) &= S(u_{i-1}\ U_{i-1}\ u_i\ U_i\ u_{i+1}\ U_{i+1}\ \dots\ u_j\ U_j)\\ &= S(X\ X\ u_{i-1}\ U_{i-1}\ u_i\ U_i\ u_{i+1}\ U_{i+1}\ \dots\ u_j\ U_j\ Y\ Y)\\ &= S(X\ u_{i-1}\ U_{i-1}\ u_i\ u_{i+1}\ \dots\ u_j\ Y)\\ &\quad \oplus S(X\ U_i\ U_{i+1}\ \dots\ U_j\ Y)\\ &= S(X\ n\ U_i\ n\ U_{i+1}\ n\ \dots\ n\ U_j\ n\ Y) \neq (0,0,\dots,0).\end{aligned}$$

Рассмотрим случай (IXb). Очевидно, что

$$S(u_m\ n+1\ U_m\ n+2\ n) = (1,\ 1,\ \dots\ ,1) \tag{12}$$

для любого $m \in N'_{l(n)}$. Тогда по доказанному выше получаем

$$\begin{aligned}(0,0,\dots,0) &= S(M)\\ &= S(X\, n+2\, n\, V_i\, n \dots n\, V_j\, n\, n+1\, U_{j+1}\, n+2\, u_{j+1}\, n+1\, Y)\\ &= S(X\, n+2\, u_{i-1}\, n+1\, U_{i-1}\, n+2\, n\, V_i\, n \dots n\, V_j\, n\, n+1\, Y)\\ &\neq (0,0,\dots,0).\end{aligned}$$

Случаи (IXc) и (IXd) сводятся к рассмотренным случаям (IXa) и (IXb) соответственно сдвигом на одну букву.

Рассмотрим случай (IXe). Из (12) имеем

$$\begin{aligned}(0,0,\dots,0) &= S(M) = S(M) \oplus (1,1,\dots,1) \oplus (1,1,\dots,1)\\ &= S(n+1\ U_{i-1}\ n+2\ n\ V_i\ n \dots\ n\ V_j\ n\ n+1\ U_{j+1}\ n+2)\\ &\oplus S(n\ n+1\ U_{i-1}\ n+2\ u_{i-1}) \oplus S(u_{j+1}\ n+1\ U_{j+1}\ n+2\ n)\\ &= S(n\ V_{i-1}\ n\ V_i\ n\ \dots\ n\ V_j\ n\ V_{j+1}\ n) \neq (0,0,\dots,0).\end{aligned}$$

Аналогично в случае (IXf) имеем

$$(0,0,\dots,0) = S(M) = S(V_i\ n\ V_{i+1}\ n\ \dots\ n\ V_j) \neq (0,0,\dots,0). \tag{13}$$

Таким образом, слово $H_{n+2} \in L^*_{n+2,n+1}$. Теорема 2 доказана.

Легко видеть, что построенное выше слово $H_n = \langle X, Y\rangle$, где X — слово в алфавите из $(n+1)/2$ букв, а Y — слово в алфавите из $(n-1)/2$ букв, причем эти алфавиты не пересекаются. Поэтому $\delta(H_n) = 1$, т. е. верно

Следствие 1. *При любом нечетном* $n \geqslant 3$

$$d(Q_n) = (n-1)2^{(n+1)/2}.$$

Из теоремы 2 с помощью простой конструкции получается следующая

Теорема 3. *При любом натуральном* $n \geqslant 3$

$$d(L_{n,n-1}) \geqslant (n-1)2^{\lceil n/2 \rceil} - 1.$$

Доказательство. Для нечетного n это непосредственно следует из теоремы 2.

Пусть n — четное. Рассмотрим следующую конструкцию. Пусть $X \in L_{n,n+1}$ и $X = Y_1\ Y_2\ \ldots\ Y_m$, где $|Y_i| = n-1$ при $i < m$, а $0 \leqslant |Y_m| \leqslant n-2$. Тогда слово $Y_1\ n+1\ Y_2\ n+1\ \ldots\ n+1\ Y_m$ принадлежит множеству $L_{n+1,n}$. Отсюда следует нужная оценка для четного n.

ЛИТЕРАТУРА

1. **Евдокимов А. А.** Метрические свойства вложений и коды, сохраняющие расстояния // Модели и методы оптимизации. Новосибирск: Наука, 1988. С. 116–132. (Тр. / АН СССР. Сиб. отд-ние. Ин-т математики; Т. 10).
2. **Евдокимов А. А.** Локально изометрические вложения графов и свойство продолжения метрики // Сиб. журн. исслед. операций. 1994. Т. 1, № 1. С. 5–12.
3. **Евдокимов А. А.** О нумерации подмножеств конечного множества // Методы дискретного анализа в решении комбинаторных задач: Сб. науч. тр. Новосибирск: Ин-т математики СО АН СССР, 1980. Вып. 34. С. 8–26.

Адрес автора:

Россия,
630090 Новосибирск,
ул. Пирогова, 2,
Новосибирский
государственный университет

Статья поступила
9 января 1996 г.