ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Июль—сентябрь 1999. Серия 1. Том 6, № 3, 3–9

УДК 519.714

# НИЖНИЕ ОЦЕНКИ МУЛЬТИПЛИКАТИВНОЙ СЛОЖНОСТИ СИМВОЛЬНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ, ОПРЕДЕЛЯЕМЫХ МОНОТОННЫМИ СИММЕТРИЧЕСКИМИ БУЛЕВЫМИ ФУНКЦИЯМИ*)

*Ю. В. Мерекин*

Для мультипликативной сложности символьных последовательностей, определяемых монотонными симметрическими булевыми функциями, в классе схем конкатенации слов получена нижняя оценка, асимптотически совпадающая с известной верхней оценкой.

## Введение

Мы рассматриваем итеративную процедуру построения слов, используя, как и в [1–3], операцию конкатенации двух слов, полученных на предыдущих шагах. В настоящей работе для одного класса слов получена нижняя оценка сложности, асимптотически совпадающая с известной верхней оценкой [3].

Рассматриваются слова в алфавите {0,1}. *Длиной* слова называется число входящих в него символов. Операция *конкатенации* слов $U$ и $V$ определяется как запись слова $V$ за словом $U$ и обозначается через $U \bullet V$. В некоторых случаях знак $\bullet$ опускается. Слово $V$ называется *подсловом* слова $W$ и обозначается через $V \subseteq W$, если для некоторых (возможно, пустых) слов $X$ и $Y$ справедливо равенство $W = X \bullet V \bullet Y$.

Последовательность слов $0, 1, X, Y, \ldots, Z$ называется *схемой конкатенации* слова $Z$ и обозначается через $S(Z)$, если для любого слова $W$ из этой последовательности, начиная со слова $X$, в ней имеются такие слова $U, V$ (возможно, $U = V$), предшествующие слову $W$, что $W = U \bullet V$. Под *сложностью* $L(S(Z))$ *схемы* $S(Z)$ понимается число слов в последовательности $X, Y, \ldots, Z$. Пусть $L(Z) = \min L(S(Z))$, где минимум берется

*) Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 99–01–00531).

по всевозможным схемам конкатенации слова $Z$. Величина $L(Z)$ называется *мультипликативной сложностью* слова $Z$.

Пусть слово $W$ представлено в виде $W = UxV$, где $x \in \{0,1\}$. Если $V$ является либо символом, отсутствующим в слове $Ux$, либо $V \subseteq Ux$, а $xV \not\subseteq U$, то слово $V$ называется *максимальным суффиксом* слова $W$ (однобуквенное слово является максимальным суффиксом). Например, для слова 00001 суффикс 1 — максимальный; для слова 010110 суффикс 10 — максимальный. Представление слова $Z$ в виде $Z = Y_1 \bullet Y_2 \bullet \ldots \bullet Y_r$ называется *суффиксным представлением,* если длина слова $Y_1$ равна единице, а всякое слово $Y_i$, $1 \leqslant i \leqslant r$, является максимальным суффиксом слова $Y_1 \bullet Y_2 \bullet \ldots \bullet Y_i$. Например, слово 00001 имеет суффиксное представление $0 \bullet 0 \bullet 00 \bullet 1$; слово 010110 имеет суффиксное представление $0 \bullet 1 \bullet 01 \bullet 10$. Очевидно, что суффиксное представление любого слова единственно. Число операций конкатенации в суффиксном представлении слова $Z$ называется *суффиксной сложностью* слова $Z$ и обозначается через $L^*(Z)$. Справедливо

**Предложение 1.** *Для любых слов $U$ и $V$ выполняется неравенство*

$$L^*(UV) \geqslant L^*(U).$$

Заметим, что операция выделения максимального суффикса по существу совпадает с так называемой операцией простого копирования, используемой в задачах сжатия информации. Поэтому суффиксная сложность имеет прямую связь с оценками сложности, полученными А. Лемпелем и Я. Зивом (см. [4–6]).

В [1] доказано, что для произвольного слова $Z$ выполняется неравенство

$$L(Z) \geqslant L^*(Z). \tag{1}$$

При нахождении суффиксного представления могут возникнуть трудности выделения максимальных суффиксов. Если в процессе построения суффиксного представления допустить замену максимального суффикса на произвольный суффикс большей длины, то можно получить некоторое разложение слова, в котором число суффиксов не превосходит их числа в суффиксном представлении и, следовательно, может использоваться при получении нижней оценки для мультипликативной сложности слова $Z$.

Пусть слово $W$ представлено в виде $W = UV$, где длина слова $V$ больше единицы. Если $V \not\subseteq U$, то слово $V$ называется *расширенным суффиксом* слова $W$ (при пустом $U$ слово $V$ является расширенным суффиксом). Например, для слова 00001 суффиксы 01, 001, 0001 и 00001 являются расширенными; для слова 010110 такими суффиксами являются 110, 0110, 10110 и 010110. Представление слова $Z$ в виде $Z = Y_1 \bullet Y_2 \ldots \bullet Y_m$

называется *расширенным представлением* и обозначается через $S^{**}(Z)$, если

(a) каждый суффикс $Y_1, Y_2, \ldots, Y_m$ является либо максимальным, либо расширенным;

(b) среди $Y_1, Y_2, \ldots, Y_m$ содержится хотя бы один расширенный суффикс.

Например, для слова 00001 представления $00 \bullet 00 \bullet 1$ и $0 \bullet 000 \bullet 1$ являются расширенными; для слова 010110 такими представлениями являются $01 \bullet 01 \bullet 10$, $0 \bullet 1 \bullet 0 \bullet 110$ и $0 \bullet 10 \bullet 110$. Расширенное представление слова, вообще говоря, не единственно. Число операций конкатенации в некотором расширенном представлении $S_i^{**}(Z)$ называется *расширенной сложностью* данного представления и обозначается через $L_i^{**}(Z)$. Очевидно

**Предложение 2.** *Для произвольного слова $Z$ и любого его расширенного представления $S_i^{**}(Z)$ выполняется неравенство*

$$L^*(Z) \geqslant L_i^{**}(Z).$$

Для слов, задаваемых столбцами значений таблиц истинности монотонных симметрических булевых функций (при лексикографическом порядке наборов значений аргументов), в классе схем конкатенации слов получим нижнюю оценку их мультипликативной сложности. Обозначим через $W_k^n$, $1 \leqslant k \leqslant n$, слово длины $2^n$, задаваемое столбцом значений таблицы истинности монотонной симметрической булевой функции от $n$ переменных, принимающей единичные значения на наборах, содержащих не менее $k$ единиц. Непосредственно из определения слова $W_k^n$ следует

**Предложение 3.** *При любых $n \geqslant 2$ и $k$, $1 \leqslant k < n$,*

$$W_k^n = W_k^{n-1} \bullet W_{k-1}^{n-1}.$$

**Лемма 1.** *В слове $W_k^n$, $1 \leqslant k < n$, имеется только одна максимальная серия из $2^{n-k+1} - 1$ единиц.*

Доказательство проведем индукцией по $n-k$. Докажем справедливость леммы для $n - k = 1$. При $n = 2$ имеем $W_1^2 = 0111$. Если $n \geqslant 3$, то применим предложение 3 сначала к слову $W_{n-1}^n$, затем — к каждому суффиксу $W_{i-1}^i$, $3 < i \leqslant n-1$, из очередного разложения. В результате получим

$$W_{n-1}^n = W_{n-1}^{n-1} \bullet W_{n-2}^{n-1} = \ldots = W_{n-1}^{n-1} \bullet W_{n-2}^{n-2} \bullet \ldots \bullet W_2^2 \bullet W_1^2, \qquad (2)$$

где $W_i^i = (0^{2^i-1})1$ для всех $i$, $2 \leqslant i \leqslant n-1$, и $W_1^2 = 0111$. Поэтому слово $W_{n-1}^n$ содержит только одну максимальную серию из трех единиц. Для $n - k = 1$ лемма доказана.

Предположим, что лемма верна при любом $t \leqslant n-k-1$. Убедимся в ее справедливости при $t = n-k$. Аналогично (2) получаем разложение слова $W_k^n$

$$W_k^n = W_k^{n-1} \bullet W_{k-1}^{n-1} = \ldots = W_k^{n-1} \bullet W_{k-1}^{n-2} \bullet \ldots \bullet W_2^{n-k+1} \bullet W_1^{n-k+1}, \quad (3)$$

в котором каждое слово $W_k^{n-1}$, $W_{k-1}^{n-2}, \ldots, W_2^{n-k+1}$ начинается с символа 0 и по индукционному предположению содержит единственную максимальную серию из $2^{n-k}-1$ единиц, а $W_1^{n-k+1} = 0(1^{2^{n-k+1}-1})$. Поэтому $W_k^n$ содержит только одну максимальную серию из $2^{n-k+1}-1$ единиц. Лемма 1 доказана.

**Лемма 2.** *Слово $W_k^n$, $2 \leqslant k < n$, представимо в виде*

$$W_k^n = W_k^k \bullet X \bullet W_1^{n-k+1}.$$

Доказательство. Применим предложение 3 сначала к слову $W_k^n$, затем еще $n-k-1$ раз к каждому префиксу $W_k^{n-1}, W_k^{n-2}, \ldots, W_k^{k+1}$ из очередного полученного разложения. Продолжим разложение, применяя $k-2$ раза предложение 3 к каждому суффиксу $W_{k-1}^{n-1}, W_{k-2}^{n-2}, \ldots, W_2^{n-(k-2)}$. В результате получаем

$$\begin{aligned} W_k^n &= (W_k^{n-1} \bullet W_{k-1}^{n-1}) = (W_k^{n-2} \bullet W_{k-1}^{n-2}) \bullet W_{k-1}^{n-1} = \ldots \\ &= (W_k^k \bullet W_{k-1}^k) \bullet \ldots \bullet W_{k-1}^{n-1} = W_k^k \bullet \ldots \bullet (W_{k-1}^{n-2} \bullet W_{k-2}^{n-2}) = \ldots \\ &= W_k^k \bullet \ldots \bullet (W_2^{n-(k-1)} \bullet W_1^{n-(k-1)}) = W_k^k \bullet X \bullet W_1^{n-k+1}. \end{aligned}$$

Лемма 2 доказана.

**Лемма 3.** *При любом $n \geqslant 3$ справедливо равенство*

$$L^*(W_{n-1}^n) = 2n.$$

Доказательство. Используя (2), зададим $W_{n-1}^n$ в виде

$$W_{n-1}^n = (0\bullet 0\bullet 0^{2^1}\bullet\ldots\bullet 0^{2^{n-3}}\bullet 0^{2^{n-2}-1}\bullet 1)\bullet(0^{2^{n-2}-1}1)\bullet\ldots\bullet(0^3 1)\bullet(01\bullet 1\bullet 1). \quad (4)$$

Представление (4) очевидно суффиксное и его сложность равна $2n$. Лемма 3 доказана.

**Лемма 4.** *Для любого слова $W_k^n$, $2 \leqslant k \leqslant n-2$, существует такое расширенное представление $S_i^{**}(W_k^n)$, что*

$$L_i^{**}(W_k^n) \geqslant L^*(W_k^{n-1}) + k.$$

Доказательство. Задав слово $W_k^n$ в виде (3) и применив для его подслов $W_k^{n-1}$, $W_{k-1}^{n-2}, \ldots, W_2^{n-k+1}$ лемму 2, получим выражение

$$\begin{aligned} W_k^n = (W_k^k X_k W_1^{n-k}) \bullet (W_{k-1}^{k-1} X_{k-1} W_1^{n-k}) \bullet \ldots \\ \bullet (W_2^2 X_2 W_1^{n-k}) \bullet (W_1^{n-k+1}), \quad (5) \end{aligned}$$

в котором все заключенные в скобки слова начинаются с символа 0 и заканчиваются либо подсловом $W_1^{n-k} = 0(1^{2^{n-k}-1})$, либо подсловом $W_1^{n-k+1} = 0(1^{2^{n-k+1}-1})$. В выражении (5) каждое заключенное в скобки слово согласно лемме 1 содержит только одну максимальную серию из $2^{n-k}-1$ единиц (для первых $k-1$ слов) и одну максимальную серию из $2^{n-k+1}-1$ единиц (для последнего слова). Поэтому суффиксное представление слова $W_k^n$ заканчивается двумя максимальными суффиксами вида $1^{2^{n-k}-1}$. Удалим из слова $W_k^n$ два полученных максимальных суффикса, которые входят в подслово $W_1^{n-k+1} = W_1^1 \bullet 1^{2^{n-k}-1} \bullet 1^{2^{n-k}-1}$, и перегруппируем подслова в выражении (5) следующим образом:

$$(W_k^k X_k W_1^{n-k} W_{k-1}^{k-1}) \bullet (X_{k-1} W_1^{n-k} W_{k-2}^{k-2}) \bullet \ldots \bullet (X_2 W_1^{n-k} W_1^1). \tag{6}$$

Покажем, что в последних $k-2$ скобках из (6) слова являются расширенными суффиксами. Выделим в (6) слова вида

$$W_1^{n-k} W_i^i = 0(1^{2^{n-k}-1})(0^{2^i-1})1, \quad 1 \leqslant i \leqslant k-1. \tag{7}$$

Легко видеть, что ни одно из $k-1$ слов (7) не является подсловом другого, а все максимальные серии из $2^{n-k}-1$ единиц в (6) принадлежат выделенным словам (7). Следовательно, в слове (6) любое подслово вида (7) единственно и $k-2$ подслов $(X_{k-1} W_1^{n-k} W_{k-2}^{k-2}), \ldots, (X_2 W_1^{n-k} W_1^1)$ являются расширенными суффиксами. Поэтому для слова $W_k^n$, в разложении которого выделено два максимальных и $k-2$ расширенных суффиксов, существует расширенное представление $S_i^{**}(W_k^n)$ такое, что

$$L_i^{**}(W_k^n) = L^*(W_k^k X_k W_1^{n-k} W_{k-1}^{k-1}) + k. \tag{8}$$

Согласно предложению 1 и лемме 2 имеем

$$L^*(W_k^k X_k W_1^{n-k} W_{k-1}^{k-1}) \geqslant L^*(W_k^k X_k W_1^{n-k}) = L^*(W_k^{n-1}).$$

Заменив в (8) величину $L^*(W_k^k X_k W_1^{n-k} W_{k-1}^{k-1})$ на $L^*(W_k^{n-1})$, получим утверждение леммы 4.

**Теорема.** *При любых $n \geqslant 3$ и $k$, $2 \leqslant k \leqslant n-1$, справедливо неравенство*

$$L(W_k^n) \geqslant k(n-k+1) + 2. \tag{9}$$

Доказательство. Для $k = n-1$ справедливость (9) следует из (1) и леммы 3. Пусть $k < n-1$. При каждом $s$, $n \geqslant s \geqslant k+2$, согласно лемме 4 имеем

$$L_i^{**}(W_k^s) + k(n-s) \geqslant L^*(W_k^{s-1}) + k(n-s+1),$$

а по предложению 2

$$L^*(W_k^{s-1}) + k(n-s+1) \geqslant L_j^{**}(W_k^{s-1}) + k(n-s+1).$$

Пользуясь этими неравенствами и чередуя применение леммы 4 и предложения 2 (начиная с леммы 4) к словам $W_k^n, W_k^{n-1}, \ldots, W_k^{k+2}$, получаем

$$L_i^{**}(W_k^n) \geqslant L^*(W_k^{k+1}) + k(n-k-1).$$

По лемме 3 имеем $L^*(W_k^{k+1}) = 2(k+1)$, что приводит к неравенству

$$L_i^{**}(W_k^n) \geqslant k(n-k+1) + 2. \tag{10}$$

Согласно неравенству (1), предложению 2 и неравенству (10) выполняются неравенства $L(W_k^n) \geqslant L^*(W_k^n)) \geqslant L_i^{**}(W_k^n) \geqslant k(n-k+1)+2$. Теорема доказана.

Нижние оценки мультипликативной сложности $L(W_1^n) = L(W_n^n) \geqslant n+1$, $n \geqslant 2$, не учтенные в теореме, легко находятся из суффиксных представлений

$$W_1^n = 0 \bullet 1 \bullet 1 \bullet 1^{2^1} \bullet \ldots \bullet 1^{2^{n-2}} \bullet 1^{2^{n-1}-1},$$
$$W_n^n = 0 \bullet 0 \bullet 0^{2^1} \bullet \ldots \bullet 0^{2^{n-2}} \bullet 0^{2^{n-1}-1} \bullet 1.$$

Полученная нижняя оценка (9) мультипликативной сложности слов, определяемых монотонными симметрическими булевыми функциями, асимптотически совпадает с верхней оценкой

$$L(W_k^n) \leqslant (k+1)(n-k+2) - 3,$$

полученной в работе [3].

Обозначим через $E_k^n$, $0 \leqslant k \leqslant n$, слово длины $2^n$, задаваемое столбцом значений таблицы истинности (при лексикографическом порядке наборов значений аргументов) элементарной симметрической булевой функции от $n$ переменных, принимающей единичные значения на наборах, содержащих ровно $k$ единиц.

Аналогично тому, как это делалось для $W_k^n$, для слов $E_k^n$ можно получить нижние оценки мультипликативной сложности

$$\begin{aligned} &L(E_0^n) \geqslant n+1 && \text{при} \quad n \geqslant 2,\\ &L(E_1^n) \geqslant 2n && \text{при} \quad n \geqslant 3,\\ &L(E_k^n) \geqslant k(n-k+1) && \text{при} \quad n \geqslant k \geqslant 2. \end{aligned}$$

Заметим, что полученные в [3] верхние оценки

$$\begin{aligned} &L(E_k^n) \leqslant (k+1)(n-k+2) - 3 && \text{при} \quad n \leqslant 2k,\\ &L(E_k^n) \leqslant (k+2)(n-k+2) - k - 5 && \text{при} \quad n \geqslant 2k, \end{aligned}$$

асимптотически совпадают с нижними.

## ЛИТЕРАТУРА

1. **Мерекин Ю. В.** Нижняя оценка сложности для схем конкатенации слов // Дискрет. анализ и исслед. операций. 1996. Т. 3, № 1. С. 52–56.
2. **Мерекин Ю. В.** О сложности символьных последовательностей, определяемых линейными булевыми функциями // Сиб. журн. индустриальной математики. 1998. Т. 1, № 1. С. 145–147.
3. **Мерекин Ю. В.** Верхние оценки сложности символьных последовательностей, порождаемых симметрическими булевыми функциями // Дискрет. анализ и исслед. операций. 1998. Т. 5, № 3. С. 38–43.
4. **Lempel A., Ziv J.** On the compexity of finite sequences // IEEE Trans. Inform. Theory. 1976. V. 22, N 1. P. 75–81.
5. **Ziv J., Lempel A.** An universal algorithm for sequential data compression // IEEE Trans. Inform. Theory. 1977. V. 23, N 3. P. 337–343.
6. **Ziv J., Lempel A.** Compression of individual sequences via variable-length coding // IEEE Trans. Inform. Theory. 1978. V. 24, N 5. P. 530–536.

Адрес автора:

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

Статья поступила
10 марта 1999 г.