СИБИРСКИЙ ЖУРНАЛ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ
Апрель – июнь 1994. Том 1, № 2, 3–7

УДК 519.114

ЧИСЛО РАЗЛИЧНЫХ ПОДСЛОВ ЗАДАННОЙ ДЛИНЫ В ПОСЛЕДОВАТЕЛЬНОСТИ МОРСА — ХЕДЛУНДА*)

С. В. Августинович

В статье получена точная формула для числа различных подслов длины n в последовательности Морса — Хедлунда [1], начальным членом которой является символ 0, а построение последующих осуществляется неограниченным применением операции замены 0 на 01 и 1 на 10. Ранее были известны лишь оценки для числа подслов заданной длины [2, 3]. Последовательность Морса — Хедлунда является классическим примером последовательности, в которой не встречаются три подряд расположенных одинаковых подслова [4].

Наш интерес к бесповторным последовательностям объясняется их связями с вопросами полноты множества слов и исследованием языков с запрещенными подсловами (см. [5]). Известно много эквивалентных способов задания последовательности Морса — Хедлунда, причем наиболее простым является следующий: i-й член последовательности есть 0, если число единиц в двоичной записи числа i четно, и равен 1 в противном случае. Другой способ — итеративный: $x_0 = 0$, $x_{2i} = x_i$, $x_{2i+1} = x_i + 1 \pmod 2$, $i = 0, 1, \ldots$.

В работе используется третий способ, наиболее удобный для рассмотрения. Пусть отображение φ переводит символ 0 в слово 01, а символ 1 в слово 10. Для произвольного слова $S = s_1 \ldots s_n$ в алфавите $\{0, 1\}$ определим $\varphi(S) = \varphi(s_1) \ldots \varphi(s_n)$. Рассмотрим семейство слов 0, 01, 0110, 01101001, ... , каждое из которых получается из предыдущего применением отображения φ и, как легко видеть, является началом следующего слова. Таким образом задается бесконечное слово $W = 0110100110010110\ldots$, которое называется последовательностью Морса — Хедлунда (см. [1]). Обозначим через $\mathfrak{M}(n)$ множество различных подслов длины n в W, а через $R(n)$ их число. А. Т. Колотовым (см. [4, 5]) были получены следующие оценки для $R(n)$:

$$2n \leqslant R(n) \leqslant 6n.$$

*) Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 93-011-1484).

В данной работе получена точная формула для $R(n)$, из которой, в частности, вытекают следующие оценки:

$$3(n-1) \leqslant R(n) \leqslant 10(n-1)/3. \tag{1}$$

Нижняя оценка достигается при $n = 2^k+1$, а верхняя — при $n = 3\cdot 2^k+1$, $k = 0, 1, \dots$.

1. Определения и обозначения

- Слово S назовем *допустимым*, если оно встречается в W в качестве подслова. Допустимое слово S назовем *бинарным справа (слева)*, если $S0$ и $S1$ ($0S$ и $1S$) допустимы. В противном случае слово S будем называть *унарным справа (слева)*.
- Упорядоченную пару (i, j), где i, $j \in \{1, 2\}$, будем называть *типом слова* S, если S i-нарно слева и j-нарно справа.

Например, слово типа (1,2) унарно слева и бинарно справа.

- Слово типа (2,2) назовем
  - — *устойчивым*, если после приписывания к нему слева любого символа получается слово бинарное справа,
  - — *неустойчивым*, если после приписывания к нему слева любого символа получается слово унарное справа.
- Слова 01 и 10 назовем *блоками*.

Из определений следует, что W можно разбить на блоки. Следовательно, и всякое допустимое слово можно разбить на блоки, за исключением, быть может, самого левого и (или) самого правого символов, которые могут оказаться обособленными. Такое разбиение слова будем называть *правильным*.

Пусть $B(n)$ обозначает число бинарных справа (в дальнейшем просто бинарных) слов длины n, а $U(n)$ и $N(n)$ — соответственно числа устойчивых и неустойчивых слов длины n.

2. Точная формула для $R(n)$

Предложение 1. *Для любого n, $n > 1$, справедливо соотношение*

$$R(n) = R(n-1) + B(n-1).$$

Доказательство. Указанное равенство отражает тот факт, что $\mathfrak{M}(n)$ получается приписыванием справа допустимых символов к словам из $\mathfrak{M}(n-1)$. При этом бинарные слова дают по дополнительному варианту каждое.

Поскольку $R(1) = 2$, справедливо

Следствие 1. *При любом $n \geqslant 2$ имеет место равенство*

$$R(n) = 2 + \sum_{i=1}^{n-1} B(i).$$

Предложение 2. *Слово типа (1,2) после приписывания к нему слева допустимого символа переходит в бинарное слово.*

Доказательство вытекает непосредственно из определений.

Предложение 3. *Среди допустимых слов длины 2 только 01 и 10 имеют тип (2,2), причем эти слова являются устойчивыми.*

Предложение 4. *Среди допустимых слов длины 3 только 010 и 101 имеют тип (2,2), причем оба слова неустойчивы.*

Справедливость утверждений 3 и 4 устанавливается непосредственно проверкой.

Предложение 5. *Все слова типа (2,2) подразделяются на устойчивые и неустойчивые, причем первые порождаются применением операции φ к словам 01 и 10, а вторые — к словам 010 и 101.*

Доказательство. Рассмотрим произвольное слово S типа (2,2) и длины $l \geqslant 5$. Легко проверить, что в слове S непременно встретятся два одинаковых символа рядом. Поэтому правильное разбиение S единственно. Если бы при таком разбиении S на каком-то из его концов оставался изолированный символ, он бы однозначно дополнялся до целого блока, что противоречит бинарности S. Значит, слово S разбивается на полные блоки, и, следовательно, существует слово S' такое, что $\varphi(S') = S$.

Рассмотрим какое-либо вхождение слова S в слово W, правильно разбитое на блоки. По определению $W = \varphi(W)$, что задает однозначное соответствие между символами слова W и блоками, на которые разбито это же слово. Из однозначности разбиения S на блоки сразу следует, что любому вхождению слова S в слово $\varphi(W)$ соответствует вхождение слова S' в слово W. То же верно и для любого допустимого расширения слова S блоками, чему будет соответствовать расширение слова S' символами. Таким образом S', как и S, допустимо и имеет тип (2,2), причем S и S' устойчивы или неустойчивы одновременно. К слову S', которое вдвое короче S, применимы аналогичные рассуждения, поэтому когда-нибудь мы придем к слову длины 3 или 4. Но среди слов длины 4 только слова 0110 и 1001 имеют тип (2,2), причем $0110 = \varphi(01)$ и $1001 = \varphi(10)$. Предложение 5 доказано.

Учитывая, что применение операции каждый раз удваивает длину слова, получим

Следствие 2. *Справедливы следующие соотношения*:

$$U(n) = \begin{cases} 2, & \text{если } n = 2^k, \ \ k = 1, 2, \dots, \\ 0 & \text{в остальных случаях;} \end{cases}$$

$$N(n) = \begin{cases} 2, & \text{если } n = 3 \cdot 2^k, \ \ k = 0, 1, \dots, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Предложение 6. *При любом* $n \geqslant 2$ *справедливо равенство*

$$B(n) = B(n-1) + U(n-1) - N(n-1).$$

Доказательство. Ясно, что все бинарные слова длины n могут быть получены из бинарных слов длины $n-1$ приписыванием слева допустимых символов. При этом согласно предложению 2 слова типа (1,2) должны быть учтены по одному разу, устойчивые слова — по два раза, а неустойчивые не должны быть учтены вообще. Предложение 6 доказано.

Теорема. *Пусть* $C(n) = 2^k$, *где* n *и* k *— натуральные числа,* $n/2 < 2^k \leqslant n$, $\rho(n) = \min\{n - C(n), 2C(n) - n\}$. *Тогда*

$$R(n+1) = 3n + \rho(n).$$

Доказательство. Из следствия 1 и предложения 6 вытекает, что

$$R(n+1) = 2 + \sum_{i=1}^{n} B(i) = 2 + \sum_{i=2}^{n} \Big(\sum_{j=2}^{i} (U(j) - N(j)) + 2 \Big)$$
$$= 2n + \sum_{i=2}^{n} \sum_{j=2}^{i} (U(j) - N(j)).$$

Пользуясь следствием 2, получаем

$$\sum_{j=2}^{i} (U(j) - N(j)) = \begin{cases} 2, & \text{если } 2^k \leqslant i < 3 \cdot 2^{k-1}, \ k = 1, 2, \dots, \\ 0, & \text{если } 3 \cdot 2^{k-1} \leqslant i < 2^{k+1}, \ k = 1, 2, \dots. \end{cases}$$

Возможны следующие случаи.

Случай 1: $2^k \leqslant n < 3 \cdot 2^{k-1}$. Тогда $C(n) = 2^k$, $\rho(n) = n - 2^k$ и

$$R(n+1) = 2n + \sum_{i=2}^{2^k} \sum_{j=2}^{i} (U(j) - N(j)) + \sum_{i=2^k+1}^{n} \sum_{j=2}^{i} (U(j) - N(j))$$
$$= 2n + 2^k + 2(n - 2^k) = 4n - 2^k = 3n + \rho(n).$$

Случай 2: $3 \cdot 2^{k-1} \leqslant n < 2^{k+1}$. Тогда $C(n) = 2^k$, $\rho(n) = 2^{k+1} - n$, и аналогично случаю 1 имеем

$$R(n+1) = 2n + \sum_{i=2}^{3 \cdot 2^{k-1}} \sum_{j=2}^{i} (U(j) - N(j)) + \sum_{i=3\cdot 2^k+1}^{n} (U(j) - N(j))$$
$$= 2n + (2^k + 2 \cdot 2^{k-1}) + 0 = 2n + 2C(n) = 3n + \rho(n).$$

Теорема доказана.

ЛИТЕРАТУРА

1. **Morse M., Hedlund G.** Symbolic dynamics // Amer. J. Math. 1936. V. 60, N 4. P. 815–866.
2. **Колотов А. Т.** Алгебры и полугруппы с квадратичными функциями роста // Алгебра и логика. 1980. Т. 19, №6. С. 659–668.
3. **Колотов А. Т.** Апериодические последовательности и функции роста алгебр // Алгебра и логика. 1981. Т. 20, №2. С. 138–154.
4. **Яглом А. М., Яглом И. М.** Неэлементарные задачи в элементарном изложении. М.: Гостехиздат, 1954.
5. **Евдокимов А. А.** Полные множества слов и их числовые характеристики // Методы дискретного анализа в исследовании экстремальных структур. Новосибирск, 1983. Вып. 39. С. 7–19.

Адрес автора:

РОССИЯ,
630090, Новосибирск,
Университетский пр., 4,
Институт математики СО РАН

Статья поступила

20 февраля 1994 г.