ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь—март 1997. Том 4, № 1, 53–59

УДК 519.114

О КОМБИНАТОРНОЙ СЛОЖНОСТИ ИТЕРАТИВНО ПОРОЖДАЕМЫХ СИМВОЛЬНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ*)

А. Э. Фрид

Получена формула для комбинаторной сложности некоторых последовательностей, построенных следующим образом: начальным символом последовательности является 0, а построение последующих символов осуществляется неограниченным применением операции замены символов $0, 1, \ldots, q-1$ на конечные слова равной длины в q-ичном алфавите. В статье обобщается результат работы [1], в которой получена формула для комбинаторной сложности последовательности Морса — Хедлунда.

1. Основные понятия

Пусть дан алфавит $\Sigma = \{0, \ldots, q-1\}$. Отображение $h : \Sigma^* \to \Sigma^*$ называется *морфизмом* [2], если $h(xy) = h(x)h(y)$ для любых слов $x, y \in \Sigma^*$.

Рассмотрим морфизм ϕ, переводящий каждое конечное слово $s = s_1 \ldots s_k$ в алфавите Σ в слово $\phi(s) = \phi(s_1) \ldots \phi(s_k)$, где $\phi(0)$, $\phi(1), \ldots$, $\phi(q-1)$ — слова равной длины $m \geqslant 2$ в алфавите Σ. Если слово $\phi(0)$ начинается с 0, то в семействе 0, $\phi(0)$, $\phi(\phi(0)), \ldots$ каждое предыдущее слово является началом следующего. Таким образом задается последовательность W, порожденная итерациями морфизма ϕ. Слово s назовем *допустимым*, если оно встречается в W в качестве подслова. Множество всех допустимых в W слов длины n обозначим через $\mathscr{M}_W(n)$. *Комбинаторной сложностью* последовательности W называется мощность множества $\mathscr{M}_W(n)$, которую будем обозначать через $R_W(n)$.

В работе получена формула для комбинаторной сложности последовательностей, порожденных итерациями морфизмов достаточно общего вида.

---

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

Слова $\phi(t)$, $t \in \Sigma$, назовем *блоками.* Из определений следует, что последовательность W можно разбить на блоки. Следовательно, всякое допустимое слово можно разбить на блоки (возможно, не единственным образом), причем самый первый и самый последний блоки могут быть неполными. Такое разбиение назовем *правильным.*

Последовательность W назовем *хорошо блокируемой*, если существует такое число N_W (*модуль блокируемости* последовательности), что правильное разбиение любого допустимого слова длины не меньшей N_W единственно, причем неполные блоки в начале и в конце этого разбиения восстанавливаются до полных единственным образом.

2. Формула для комбинаторной сложности хорошо блокируемых последовательностей

Пусть W — хорошо блокируемая последовательность с модулем блокируемости N_W, порожденная итерациями морфизма ϕ, m — длина блока, d — натуральное число, удовлетворяющее условию $md-m+1 \geqslant N_W$. Пусть число $r \in \{d+1,\dots,md\}$ таково, что для некоторого целого неотрицательного p выполняется соотношение

$$m^p(r-1) < n \leqslant m^p r.$$

Такое r существует для любого n и определяется равенством

$$r = \left\lfloor \frac{n-1}{\lfloor (n-1)/d \rfloor} \right\rfloor + 1.$$

Пусть также $\Delta = m^p r - n$. Справедлива следующая

Теорема. *При любом $n > d$ выполняется равенство*

$$R_W(n+1) = \frac{1}{r}\{nR_W(r+1) + \Delta(rR_W(r) - (r-1)R_W(r+1))\}.$$

Доказательство. *Степенью ветвления* $c(s)$ допустимого слова s длины n назовем число таких символов t из q-ичного алфавита, что после приписывания их к s справа получается допустимое слово

$$c(s) = |\{t \mid t \in \Sigma, \quad st \in \mathscr{M}_W(n+1)\}|.$$

Через $B_j(k)$ обозначим множество допустимых слов длины k со степенью ветвления, равной j, т. е.

$$B_j(k) = \{s \mid |s| = k, \quad c(s) = j\}.$$

Положим также

$$G_W(i) = R_W(i+1) - R_W(i).$$

Предложение 1. *Для любого k выполняется равенство*

$$G_W(k) = |B_2(k)| + 2\,|B_3(k)| + \ldots + (q-1)\,|B_q(k)|. \tag{1}$$

Доказательство. Указанное равенство отражает тот факт, что $\mathscr{M}_W(k+1)$ получается приписыванием справа допустимых символов к словам из $\mathscr{M}_W(k)$. При этом из каждого слова со степенью ветвления $j \geqslant 2$ получается $j-1$ дополнительное слово. Предложение 1 доказано.

Для каждого допустимого слова s и для каждого $i \in \{0, \ldots, m-1\}$ определим слово $\psi_i(s)$, получаемое из $\phi\,(s)$ удалением первых i символов.

Предложение 2. *Для любого допустимого слова s длины не меньше d выполняется равенство*

$$c\,(s) = c\,(\phi\,(s)).$$

Доказательство. Указанное равенство следует из того соображения, что приписывание к слову s справа символов соответствует приписыванию к слову $\phi\,(s)$ справа блоков и ввиду хорошей блокируемости каждый блок однозначно определяется первым символом.

Предложение 3. *Для любого допустимого слова s длины не меньше d и для любого $i < m$ выполняется равенство*

$$c\,(\phi\,(s)) = c\,(\psi_i(s)).$$

Доказательство. Слово $\psi_i(s)$ имеет длину, не меньшую модуля блокируемости последовательности. Поэтому $\psi_i(s)$ имеет единственное правильное разбиение на блоки, соответствующее правильному разбиению слова $\phi\,(s)$. Следовательно, вхождения слова $\psi_i(s)$ встречаются в последовательности W только как части вхождений слова $\phi\,(s)$. Предложение 3 доказано.

Из предложений 2 и 3 вытекает

Следствие 1. *Для любого допустимого слова s длины d выполняется равенство*

$$c\,(\psi_i(s)) = c\,(s). \tag{2}$$

Предложение 4. *Пусть $k \leqslant d$. Тогда для всех $j \geqslant 2$*

$$|B_j(k)| = |B_j(mk)| = \ldots = |B_j(mk-m+1)|. \tag{3}$$

Доказательство. Из (2) следует, что при каждом $i \in \{0, \ldots, m-1\}$ отображение ψ_i отображает множество $B_j(k)$ на множество $B_j(mk-i)$. Докажем, что при $j \geqslant 2$ на множестве $B_j(k)$ существует отображение, обратное к ψ_i, из чего и будет следовать равенство мощностей множеств $B_j(k)$ и $B_j(mk-i)$.

Пусть v — слово длины $mk - i$ со степенью ветвления $j \geqslant 2$. Из условия хорошей блокируемости следует, что единственное правильное разбиение слова v кончается целым блоком и каждое вхождение в W слова v является частью вхождения некоторого слова uv длины mk, состоящего из k полных блоков. Существует слово v' такое, что $\phi(v') = uv$; из задания последовательности следует, что v' допустимо. По построению $\psi_i(v') = v$. Отсюда и из (2) получаем, что $c(v') = c(v) = j$. Поэтому $v = \psi_i(v')$, и из однозначности построения слова v' следует, что $v' = (\psi_i)^{-1}(v)$. Отображение ψ_i взаимно однозначно на $B_j(k)$. Предложение 4 доказано.

Из (1) многократным применением (3) получаем

Следствие 2. *Пусть $m^{p_k}(r_k - 1) < k \leqslant m^{p_k} r_k$ для некоторого целого неотрицательного p_k и $r_k \in \{d+1, \dots, md\}$. Тогда*

$$G_W(k) = G_W(r_k). \tag{4}$$

Предложение 5. *Если число k таково, что $mk + 1 \geqslant N_W$, то*

$$R_W(mk+1) = mR_W(k+1).$$

Доказательство. Положим

$$\phi(\mathscr{M}_W(k+1)) = \{\phi(s) \mid s \in \mathscr{M}_W(k+1)\}.$$

Очевидно, все слова из $\phi(\mathscr{M}_W(k+1))$ допустимы в W и $|\phi(\mathscr{M}_W(k+1))| = |\mathscr{M}_W(k+1)| = R_W(k+1)$.

Каждое слово из $\phi(\mathscr{M}_W(k+1))$ имеет длину $m(k+1)$ и содержит точно m подслов длины $mk+1$, причем все они допустимы в W и различны (так как правильное разложение слова длины не меньше $mk+1$ единственно). С другой стороны, каждое допустимое слово s_1 длины $mk+1$ из-за хорошей блокируемости имеет единственное правильное разложение на блоки и однозначно восстанавливается до слова, состоящего из $k+1$ целого блока, т. е. до слова из $\phi(\mathscr{M}_W(k+1))$. Поэтому $R_W(mk+1) = m|\phi(\mathscr{M}_W(k+1))| = mR_W(k+1)$. Предложение 5 доказано.

В применении к последовательности Морса —Хедлунда последнее предложение было доказано в [3].

Доказательство теоремы. Очевидно, что

$$\begin{aligned} R_W(n+1) &= R_W(d+1) + \sum_{i=d+1}^{n} G_W(i) \\ &= R_W(d+1) + \sum_{i=d+1}^{md} G_W(i) + \sum_{i=md+1}^{m^2 d} G_W(i) + \ldots + \sum_{i=m^p+1}^{n} G_W(i). \end{aligned}$$

Из (4) следует, что

$$\sum_{i=m^j d+1}^{m^{j+1}d} G_W(i) = m^j \sum_{i=d+1}^{md} G_W(i).$$

Поэтому

$$\begin{aligned} R_W(n+1) &= R_W(d+1) + (1+m+\ldots+m^{p-1}) \sum_{i=d+1}^{md} G_W(i) \\ &+ \sum_{i=m^p d+1}^{m^p(d+1)} G_W(i) + \ldots + \sum_{i=m^p(r-2)}^{m^p(r-1)} G_W(i) + \sum_{i=m^p(r-1)}^{n} G_W(i) \\ &= R_W(d+1) + \frac{m^p-1}{m-1}(R_W(md+1) - R_W(d+1)) \\ &\qquad + m^p \sum_{i=d+1}^{r-1} G_W(i) + (n - m^p(r-1))(G_W(r)). \end{aligned}$$

Наконец, воспользовавшись предложением 5, получаем

$$\begin{aligned} R_W(n+1) &= R_W(d+1) + (m^p-1)R_W(md+1) \\ &+ m^p(R_W(r) - R_W(d+1)) + (m^p - \Delta)(G_W(r)) \\ &= m^p R_W(r+1) - \Delta(R_W(r+1) - R_W(r)). \end{aligned}$$

Ввиду того, что $m^p = (\Delta + n)/r$, имеем

$$R_W(n+1) = \frac{n}{r} R_W(r+1) + \frac{\Delta}{r}(r R_W(r) - (r-1)R_W(r+1)).$$

Теорема доказана.

3. Оценки для комбинаторной сложности

Заметим, что все участвующие в формуле величины могут быть получены конечным перебором. Чтобы вычислить значения $R_W(n)$ для любого n, достаточно знать значения $R_W(r)$ для $r \leqslant md+1$, где числа m и d зависят только от самой последовательности.

Нетрудно видеть, что на промежутках вида

$$m^p(r-1) < n \leqslant m^p r$$

зависимость $R_W(n+1)$ от n является линейной:

$$R_W(n+1) = a(r)n + b(r,p).$$

В целом же график функции $R_W(n+1)$ представляет собой ломаную с концами отрезков в точках $n = m^p r$, где $r \in \{d+1, \dots, md\}$, т. е. в точках, соответствующих $\Delta(n) = 0$. Отсюда вытекает, что если

$$c = \min_{r\in\{d+1,\dots,md\}} \frac{R_W(r+1)}{r} = \frac{R_W(k+1)}{k},$$

$$C = \max_{r\in\{d+1,\dots,md\}} \frac{R_W(r+1)}{r} = \frac{R_W(K+1)}{K},$$

то при всех $n > d$ верны оценки

$$c \leqslant \frac{R_W(n+1)}{n} \leqslant C.$$

Эти оценки неулучшаемы и достигаются на бесконечном числе точек:

$$\frac{R_W(n+1)}{n} = c \text{ при } n = m^p k, \text{ где } p = 0, 1, \dots,$$

$$\frac{R_W(n+1)}{n} = C \text{ при } n = m^p K, \text{ где } p = 0, 1, \dots$$

4. Примеры хорошо блокируемых последовательностей

Простейшей хорошо блокируемой последовательностью является последовательность Морса — Хедлунда, задаваемая морфизмом

$$\begin{cases} \phi(0) = 01, \\ \phi(1) = 10. \end{cases}$$

Легко видеть, что уже каждое допустимое в последовательности Морса — Хедлунда слово длины 4 однозначно разбивается на блоки. Поэтому можно положить $d = 3$ и, вычислив $R_W(n)$ для $n = 3, 4, 5, 6, 7$, после простейших преобразований получить следующую формулу:

$$R_W(n+1) = \begin{cases} 2n + 2^{k+1}, & \text{если } 3 \cdot 2^{k-1} \leqslant n < 2^{k+1}, \\ 4n - 2^k, & \text{если } 2^k \leqslant n < 3 \cdot 2^{k-1}. \end{cases}$$

Этот результат получен ранее в [1].

Нетрудно показать, что хорошо блокируется всякая бинарная последовательность, порожденная итерациями морфизма ϕ вида

$$\begin{cases} \phi(0) = 0s_1 1, \\ \phi(1) = 1s_2 0, \end{cases}$$

где s_1, s_2 — произвольные бинарные слова равной длины. То же самое можно сказать о морфизмах вида

$$\begin{cases} \phi(0) = 0s_1 0, \\ \phi(1) = 1s_2 1, \end{cases}$$

за исключением следующих морфизмов:

$$1)\ \begin{cases} \phi(0) = 0101\ldots0, \\ \phi(1) = 1010\ldots1, \end{cases}$$

$$2)\ \begin{cases} \phi(0) = 00\ldots0, \\ \phi(1) = 1X1, \end{cases}$$

$$3)\ \begin{cases} \phi(1) = 11\ldots1, \\ \phi(0) = 0X0, \end{cases}$$

где X — произвольное слово в алфавите $\{0, 1\}$.

Морфизмы первых двух видов порождают простейшие периодические последовательности с комбинаторной сложностью $R_W(n) \equiv 2$ и $R_W(n) \equiv 1$ соответственно. Комбинаторная сложность последовательностей, порожденных итерациями морфизмов третьего вида, может быть вычислена с помощью метода, аналогичного описанному.

Легко показать, что не существует других хорошо блокируемых бинарных последовательностей. Более того, чтобы последовательность над алфавитом Σ обладала свойством хорошей блокируемости, очевидно, необходимо, чтобы порождающий ее морфизм ϕ имел вид

$$\phi(i-1) = \sigma(i)s_i\pi(i),$$

где σ, π — перестановки над Σ, s_i — слово длины $m-2$ над алфавитом Σ.

Автор признателен С. В. Августиновичу за внимание и доброжелательную критику на всех этапах работы.

ЛИТЕРАТУРА

1. **Августинович С. В.** Число различных подслов заданной длины в последовательности Морса — Хедлунда // Сиб. журн. исслед. операций. 1994. Т. 1, № 2. С. 3–7.
2. **Саломаа А.** Жемчужины теории формальных языков. М.: Мир, 1986.
3. **Dekking F. M.** On the Thue — Morse measure // Acta Univ. Carolin. Math. Phys. 1992. V. 33, N 2. P. 35–40.

Адрес автора:

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

Статья поступила
28 ноября 1996 г.