ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь—март 1998. Серия 1. Том 5, № 1, 82–87

УДК 519.114

ЧАСТОТА ВХОЖДЕНИЯ СЛОВ В DOL-ПОСЛЕДОВАТЕЛЬНОСТЬ*)

А. Э. Фрид

Получена рекуррентная формула для вычисления эргодической меры μ, порожденной DOL-последовательностью, т. е. символьной последовательностью, задаваемой с помощью итераций морфизма. Фактически мера $\mu(u)$ слова u представляет собой частоту, с которой слово u встречается в последовательности.

Введение

DOL-последовательности составляют хорошо изученный класс символьных последовательностей. К настоящему времени они активно исследовались как с точки зрения комбинаторики и теории формальных языков, так и с точки зрения порожденных ими топологических систем.

Эргодическая мера μ, порожденная DOL-последовательностью, может быть описана как частота, с которой каждое слово встречается в этой последовательности: для каждого слова u значение $\mu(u)$ равно пределу отношения числа вхождений слова u в начальный отрезок DOL-последовательности к длине этого начального отрезка. Условия существования меры μ подробно изучены в [6]. Там же приведен алгоритм ее вычисления.

В [3] дано полное описание эргодической меры, порожденной двумя конкретными хорошо известными примерами неподвижных точек морфизмов, а именно последовательностью Туэ–Морса и последовательностью Фибоначчи.

В данной работе впервые предложен сравнительно простой естественный подход к вычислению эргодической меры, порожденной произвольной DOL-последовательностью (если, конечно, эта мера для данной последовательности существует). Получена рекуррентная формула,

*) Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 96–01–01800) и Федеральной целевой программы «Интеграция» (код проекта 473).

позволяющая выразить меру произвольного слова, используя меру более коротких слов. Применены те же комбинаторные методы, что и в работах [1, 2, 5].

1. Основные определения и некоторые известные результаты

Пусть $\Sigma = \{1, \ldots, q\}$ — заданный алфавит. Через Σ^* обозначается множество всех конечных слов в алфавите Σ.

Отображение $\varphi : \Sigma^* \to \Sigma^*$ называется *морфизмом*, если $\varphi(uv) = \varphi(u)\varphi(v)$ для всех u, v из Σ^*. Ясно, что морфизм φ может быть задан значениями $\varphi(i)$, где $i \in \Sigma$. Слова $\varphi(i)$, $i \in \Sigma$, будем называть *блоками*. Длина слова u обозначается через $|u|$.

Пусть для некоторого $a \in \Sigma$ слово $\varphi(a)$ начинается с a, а длина слов $\varphi^k(a)$ неограниченно возрастает. *Неподвижной точкой* морфизма φ называется такая символьная последовательность $w = w(\varphi)$, что

$$w(\varphi) = \lim_{k \to \infty} \varphi^k(a).$$

Без ограничения общности считаем, что $a = 1$. Неподвижные точки морфизмов называются также DOL-*последовательностями.*

Слово v называется *подсловом* слова u, если $u = s_1 v s_2$ для некоторых $s_1, s_2 \in \Sigma^*$. Слово v называется *допустимым* (в последовательности w), если оно является подсловом какого-то начального отрезка последовательности w. Здесь и далее $L_v(u)$ обозначает число вхождений слова v в слово u в качестве подслова. Через w_n обозначается начальный отрезок длины n последовательности w.

Инвариантная относительно сдвига эргодическая мера $\mu_w = \mu$, порожденная DOL-последовательностью w, определяется для каждого слова $u \in \Sigma^*$ как предел

$$\mu(u) = \lim_{n \to \infty} \frac{L_u(w_n)}{n}. \qquad (1)$$

Вообще говоря, предел, стоящий в определении меры μ, может и не существовать. Приведенное ниже достаточное условие существования предела (1) доказано в [5].

Матрицей морфизма φ называется квадратная матрица $M(\varphi)$ порядка q, элементы которой обозначают число вхождений символов в блоки: $m_{ij} = L_i(\varphi(j))$. Заметим, что $M(\varphi^n) = (M(\varphi))^n$.

Матрица M называется *примитивной*, если для некоторого натурального k элементы матрицы M^k строго положительны. Морфизм φ называется *примитивным*, если его матрица $M(\varphi)$ примитивна.

Теорема 1 (Perron–Frobenius). *Пусть матрица M примитивна и ее компоненты неотрицательны. Тогда у матрицы M существует строго положительное собственное число θ кратности один, превосходящее по модулю любое другое собственное число матрицы M. Существует собственный вектор со строго положительными компонентами, соответствующий собственному числу θ.*

Собственное число θ называется *собственным числом Перрона–Фробениуса* матрицы M.

Следующие предложения были доказаны в [5].

Предложение 1. *Если морфизм φ примитивен, то мера μ, определенная равенством (1), существует, и если слово u встречается в последовательности $w(\varphi)$, то $\mu(u) > 0$.*

Рассмотрим эргодическую меру μ, порожденную неподвижной точкой примитивного морфизма $w(\varphi)$.

Предложение 2. *Вектор, составленный из значений $\mu(i)$, где $i \in \Sigma$, является собственным вектором матрицы $M(\varphi)$, соответствующим ее собственному значению Перрона–Фробениуса.*

Заметим, что, поскольку $\sum\limits_{i\in\Sigma} \mu(i) = 1$, предложение 2 позволяет однозначно определить $\mu(i)$ по матрице морфизма.

Предложение 3. *Существует предел θ отношения длин слов $\varphi^{n+1}(a)$ и $\varphi^n(a)$, не зависящий от начального слова a и равный собственному числу Перрона–Фробениуса матрицы морфизма φ, т. е.*

$$\lim_{n\to\infty} \frac{|\varphi^{n+1}(a)|}{|\varphi^n(a)|} = \theta.$$

Мера μ, определенная равенством (1), может существовать и в случае, когда морфизм не является примитивным. В дальнейшем нам не понадобятся ни примитивность морфизма φ, ни строгая положительность значений меры μ слов, встречающихся в $w(\varphi)$. Однако мы предполагаем, что верны следующие допущения.

1. Для любого слова $u \in \Sigma^*$ существует мера $\mu(u) = \lim\limits_{n\to\infty} \frac{L_u(w_n(\varphi))}{n}$.
2. Существует $\lim\limits_{n\to\infty} \frac{|\varphi^{n+1}(1)|}{|\varphi^n(1)|} = \theta > 1$.

Напомним, что начальный символ неподвижной точки морфизма φ мы считаем равным 1.

Из этих двух предположений выведем формулу для значений меры μ, для применения которой достаточно знать значения $\mu(i)$ для $i \in \Sigma$.

2. Формула и ее применение

Тройка $(a_0a_1\dots a_{n+1}, i, j)$, где $a_0a_1\dots a_{n+1} \in \Sigma^*$, $0 \leqslant i < |\varphi(a_0)|$ и $0 \leqslant j < |\varphi(a_{n+1})|$, называется *интерпретацией* слова u относительно морфизма φ, если u может быть получено из слова $\varphi(a_0\dots a_{n+1})$ стиранием i первых и j последних символов. Слово $a_0a_1\dots a_{n+1}$ называется *предком* слова u или, что точнее, предком его интерпретации $(a_0a_1\dots a_{n+1}, i, j)$. Заметим, что слово $a_0a_1\dots a_{n+1}$ может быть предком нескольких различных интерпретаций слова u.

Множество всех интерпретаций слова u обозначаем через $I(u)$. Предок интерпретации $s \in I(u)$ обозначается через $a(s)$.

Заметим, что слово допустимо в D0L-последовательности $w(\varphi)$ тогда и только тогда, когда хотя бы один его предок допустим в $w(\varphi)$.

Предложение 4. *При всех $u, v \in \Sigma^*$ выполняется равенство*

$$L_v(\varphi(u)) = \sum_{s\in I(v)} L_{a(s)}(u).$$

Доказательство. Каждому вхождению слова v в $\varphi(u)$ соответствует вхождение какого-то его предка в u и наоборот.

Теорема 2. *Для любого слова $u \in \Sigma$ выполняется равенство*

$$\mu(u) = \frac{1}{\theta}\sum_{s\in I(u)} \mu(a(s)). \qquad (2)$$

Доказательство. Из предложения 4 следует, что

$$L_u(\varphi^{n+1}(1)) = \sum_{s\in I(u)} L_{a(s)}(\varphi^n(1)).$$

Разделив последнее равенство на длину слова $\varphi^{n+1}(1)$ и переходя к пределу при $n \to \infty$, получаем

$$\begin{aligned}\mu(u) &= \lim_{n\to\infty}\frac{L_u(\varphi^{n+1}(1))}{|\varphi^{n+1}(1)|} = \sum_{s\in I(u)}\lim_{n\to\infty}\frac{L_{a(s)}(\varphi^n(1))}{|\varphi^{n+1}(1)|}\\ &= \sum_{s\in I(u)}\lim_{n\to\infty}\frac{|\varphi^n(1)|}{|\varphi^{n+1}(1)|}\cdot\frac{L_{a(s)}(\varphi^n(1))}{|\varphi^n(1)|}\\ &= \sum_{s\in I(u)}\frac{1}{\theta}\mu(a(s)) = \frac{1}{\theta}\sum_{s\in I(u)}\mu(a(s)).\end{aligned}$$

Формула (2) является основной для вычисления значений меры μ. Будем считать, что известны значения $\mu(i)$ для символов $i \in \Sigma$ (см. предложение 2). Если морфизм φ является *нестирающим*, т. е. никакой блок

$\varphi(i)$, где $i \in \Sigma$, не является пустым словом, то по формуле (2) можно последовательно найти значения меры μ на всех словах длины 2, 3 и т. д.

Действительно, предположим для простоты, что морфизм φ является *строго удлиняющим*, т. е. $|\varphi(i)| \geqslant 2$ для всех символов $i \in \Sigma$. В том случае, если $|u| \geqslant 3$, а слово v является предком слова u, $|v| < |u|$. Поэтому, чтобы по формуле (2) последовательно найти значения меры μ для всех слов длины больше двух, достаточно знать значения $\mu(u)$ для всех слов u длины один и два.

Итак, пусть мера каждого слова длины один известна. Рассмотрим произвольное слово u длины два и обозначим через $A(u)$ множество всех интерпретаций слова u, предки которых имеют длину два. Формула (2) для слова u преобразуется к виду

$$\mu(u) = \frac{1}{\theta}\left(c(u) + \sum_{s \in A(u)} \mu(a(s))\right), \tag{3}$$

где $c(u)$ равна сумме значений меры μ на предках длины один интерпретаций слова u. Система уравнений (3), выписанных для всех слов длины два над алфавитом Σ, представляет собой систему из q^2 линейных уравнений с q^2 неизвестными. Так как каждое слово длины два является предком только одного слова длины два, а $\theta > 1$, то строки этой системы линейно независимы. Значит, система позволяет найти значения меры μ для всех слов длины два.

После этого, как было замечено выше, формулу (2) можно применять напрямую для вычисления значения меры μ любого слова длины не меньше трех. Если же морфизм не является строго удлиняющим, то, возможно, приходится решать еще одну или несколько аналогичных систем для мер слов длины три, четыре и более, после чего формула (2) применяется напрямую.

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. **Фрид А. Э.** О комбинаторной сложности итеративно порождаемых символьных последовательностей // Дискрет. анализ и исслед. операций. Сер. 1. 1997. Т. 4, № 1. С. 53–59.

3. **Dekking M.** On the Thue–Morse measure // Acta Univ. Carolin. Math. Phys. 1992. V. 33, N 2. P. 35–40.

4. **Frid A. E.** The subword complexity of fixed points of binary uniform morphisms // Fundamentals of computation theory. Berlin: Springer, 1997. P. 179–187. (Lecture Notes in Comput. Sci.; V. 1279).
5. **Mignosi F., Séébold P.** If a D0L language is k-power-free then it is circular // Automata, languages and programming. Berlin: Springer-Verlag, 1993. P. 507–518. (Lecture Notes in Comput. Sci.; V. 700).
6. **Queffelec M.** Substitution dynamical systems — spectral analysis. Berlin: Springer-Verlag, 1987. (Lecture Notes in Math.; V. 1294).

Адрес автора:

Институт математики
им. С. Л. Соболева СО РАН,
пр. Академика Коптюга, 4,
630090 Новосибирск,
Россия

Статья поступила
6 октября 1997 г.