Мальцевские чтения
Новосибирск (2005)
Тезисы докладов

Функция сложности и области насыщения 01-слов

М. В. Тихомирова

На примере 01-слов исследуется комбинаторное строение слов. Для количественной характеристики сложности строения слова вводится функция сложности. Особое внимание уделяется точке, в которой прекращается рост функции, — точке насыщения. Исследуется множество точек насыщения всех слов фиксированной длины. Проверена гипотеза о совпадении реальной и гипотетической областей насыщения 01-слов длины $\leqslant 20$.

Понятие функции сложности является количественной характеристикой комбинаторного строения слова фиксированной длины. Впервые понятие функции сложности было введено профессором Беляевым В. В.

В ходе исследования огромного фактического материала, были выявлены закономерности роста и убывания функции сложности, общие для всех 01-слов. В связи с этим особое значение приобретает *точка насыщения* — точка, в которой рост функции прекращается. Множество всех точек насыщения образует *область насыщения* слов фиксированной длины. Было доказано [**2**], что все точки насыщения слов фиксированной длины попадают в ограниченную область, названную *гипотетической областью насыщения*. Естественно возникает вопрос (Беляев В. В.): *верно ли, что реальная и гипотетическая области совпадают?* Доказан факт совпадения областей насыщения для всех 01-слов длины $\leqslant 13$ [**2**]. Проверка гипотезы для слов большей длины требует проведения серьезной теоретической работы.

Основные понятия и некоторые результаты

Конечные слова в алфавите $\{0, 1\}$ будем называть 01-*словами*.

Слова вида: $a^n = \underbrace{aa\ \ldots\ a}_{n}$ будем называть *стационарными*. В противном случае будем говорить, что слово *нестационарное*.

Слово, не содержащее букв, будем называть *пустым* и обозначать через $\varnothing$.

Произведением слова $u = u_1u_2\ \ldots\ u_n$ *на слово* $v = v_1v_2\ \ldots\ v_n$ будем называть слово $uv = u_1u_2\ \ldots\ u_nv_1v_2\ \ldots\ v_n$.

Пусть w — некоторое слово. Слово v будем называть *подсловом* слова w, если $w = avb$, где a и b — некоторые слова (причем, a и b могут быть пустыми).

Определение 1 (Беляев). Пусть w — произвольное 01-слово длины L. Для каждого целого $n = 0, 1, \ldots, L$ определим *функцию сложности* $f_w(n)$

следующим образом: $f_w(n)$ — количество различных подслов длины n в слове w.

Для дальнейшего исследования функции сложности понадобится ввести классификацию точек дискретной функции. Пусть $f(n)$ — произвольная дискретная функция, определенная для $n = 0, 1, \dots, L$, принимающая целые значения. Целое число m из промежутка $[0; L)$ будем называть *точкой возрастания* функции $f(m)$, если $f(m+1) > f(m)$, *точкой убывания*, если $f(m+1) < f(m)$ и *стационарной точкой*, если $f(m+1) = f(m)$.

Лемма 1. *Пусть w — произвольное 01-слово длины L, $f_w(n)$ — функция сложности слова w. Тогда для любого $n \in [0; L]$*
(1) $f_w(n) \leqslant 2^n$,
(2) $f_w(n) \leqslant L + 1 - n$.

Лемма 2. *Пусть w — нестационарное 01-слово длины L, $f_w(n)$ — функция сложности слова w и пусть $f_w(n)$ убывает на промежутке $[k_0; L)$. Тогда $f_w(n) = L + 1 - n$ для всех $n \in [k_0; L)$.*

В ходе исследования большого фактического материала были выявлены закономерности поведения функции сложности нестационарных слов, сформулированные в виде следующей гипотезы.

Гипотеза 1. Пусть w — произвольное слово длины L, $f_w(n)$ — функция сложности слова w, n — произвольное целое число из $[0; L)$. Тогда
(1) если n — точка возрастания, то $n + 1$ — тоже точка возрастания или стационарная точка;
(2) если n — стационарная точка, то $n + 1$ — тоже стационарная точка или точка убывания;
(3) если n — точка убывания, то $n + 1$ — тоже точка убывания.

Дальнейшие исследования функции сложности будем проводить, полагая, что сформулированная гипотеза справедлива. В этом случае большое значение приобретает точка, в которой функция сложности перестает расти, насыщается. И если закономерности роста функции сложности пока мало изучены, то, основываясь на леммах 1 и 2, дальнейшее поведение функции можно описать следующим образом: функция стационарна до пересечения с прямой $y = L + 1 - n$, а затем убывает по прямой $y = L + 1 - n$.

Определение 2. Пусть w — непустое 01-слово длины L, $f_w(n)$ — функция сложности слова w и пусть $n_0 \in [0; L)$ такое, что
(1) $f_w(n_0) \leqslant f_w(n_0 + 1)$,
(2) $f_w(n) \geqslant f_w(n + 1)$ для всех $n \geqslant n_0$.
Тогда точку $N(n_0; f_w(n_0))$ будем называть *точкой насыщения* слова w.

Определение 3. Множество точек насыщения непустых 01-слов длины L будем называть *областью насыщения* слов длины L и обозначать через D_L.

Рассмотрим следующее множество.

Определение 4. Обозначим через D_L^* множество всех точек с целочисленными координатами, лежащих в криволинейном треугольнике, ограниченном линиями $y = 2^n$, $y = L + 1 - n$ и $y = n + 1$. Это множество D_L^* в дальнейшем будем называть *гипотетической областью насыщения* слов длины L.

Гипотеза 2 (Беляев В. В.). $D_L = D_L^*$, для любого натурального L.

На сегодняшний день доказано следующее [**2**].

Теорема 1. *$D_L \subseteq D_L^*$ для любого натурального L.*

Таким образом, гипотеза 2 редуцируется в следующую.

Гипотеза 3. $D_L \supseteq D_L^*$, для любого натурального L.

В ходе дальнейшего исследования рассматривались слова вида ua^k.

Теорема 2. *Пусть $w = va^k u$ — произвольное 01-слово, где $a \in \{0; 1\}$, $k \in N$, причем u и v не имеют подслова a^k, а w не содержит подслова a^{k+1}. Пусть $w' = va^{k+1}u$. Тогда*

$$f_{w'}(n) = \begin{cases} f_w(n), & \text{если } n \leqslant k, \\ f_w(n) + 1, & \text{если } n > k. \end{cases}$$

Лемма 3. *Пусть $w = ua^k$ — произвольное 01-слово, где $a \in \{0; 1\}$, $k \in N$, причем u не содержит подслова a^k, а w не содержит подслова a^{k+1}. Пусть k — точка убывания функции сложности слова w. Тогда точка насыщения слова w будет являться точкой насыщения слова $w = ua^{k'}$ для любого $k' \geqslant k$.*

Обозначим через D^*, множество всех точек с целочисленными координатами, лежащих в области, ограниченной линиями: $y = 2^n$ и $y = n + 1$.

Очевидно, что $D^* = \bigcup D_L^*$.

Гипотеза 4. Пусть $P(n; n+k) \in D^*$, $n \in N$, $k \leqslant 17$. Тогда точка $P(n; n+k)$ является точкой насыщения слова $w_k(n)$ (см. табл. 1).

Table 1. Общий вид слов $w_k(n)$

k	n_{min}	точка насыщения	$w_k(n)$
1	0	$(n; n+1)$	1^n0^n
2	2	$(n; n+2)$	01^n0^n
3	3	$(n; n+3)$	101^n0^n
4	3	$(n; n+4)$	0101^n0^n
5	3	$(n; n+5)$	00101^n0^n
6	4	$(n; n+6)$	100101^n0^n
7	4	$(n; n+7)$	0100101^n0^n
8	4	$(n; n+8)$	10100101^n0^n
9	4	$(n; n+9)$	110100101^n0^n
10	4	$(n; n+10)$	0110100101^n0^n
11	4	$(n; n+11)$	00110100101^n0^n
12	4	$(n; n+12)$	000110100101^n0^n
13	5	$(n; n+13)$	1000110100101^n0^n
14	5	$(n; n+14)$	01000110100101^n0^n
15	5	$(n; n+15)$	001000110100101^n0^n
16	5	$(n; n+16)$	0001000110100101^n0^n
17	5	$(n; n+17)$	00001000110100101^n0^n

Нетрудно заметить, что каждый последующий пример получен из предыдущего умножением на одну букву (0 или 1) слева. К сожалению, на $w_{17}(n)$ цепочка обрывается, и ни $0w_{17}(n)$, ни $1w_{17}(n)$ не подходят в качестве примера для класса точек $\{(n; n+18)\}$. Описанный метод построения примеров успешно применен при решении основной проблемы — проверки гипотезы 3 для $L \leqslant 26$.

Преимущество данного метода, заключается в том, что к слову $w_k(n)$ можно применить лемму 3. Таким образом, вместо одного слова, мы получаем целый класс слов, имеющих одну и ту же точку насыщения. Подчеркнем, что это бесконечный класс слов сколь угодно большой длины.

Основные результаты

Рассмотрим два множества: D^*_{L-1} и D^*_L. Очевидно, что $D^*_{L-1} \subseteq D^*_L$.

Предположение. Если $D^*_{L-1} \subseteq D_{L-1}$, то $D^*_{L-1} \subseteq D_L$.

Если предположение верно, то каждую точку достаточно исследовать лишь раз, для наименьшего значения L, при котором эта точка попадает в какую-либо гипотетическую область насыщения. В этом случае для каждого значения L достаточно рассмотреть точки, лежащие на границе множества D^*_L, задаваемой уравнением $y = L + 1 - n$, и доказать, что они являются точками насыщения слов длины $L' \geqslant L$.

Теорема 3. *Пусть $L \leqslant 26$. Тогда $D^*_L \subseteq D_{L'}$ для всех $L' \geqslant L$.*

Следствие 1. *Если $L \leqslant 26$, то $D_L = D^*_L$.*

Другими словами, при $L = 1, 2, \ldots, 26$ точки гипотетической области насыщения D^*_L принадлежат не только D_L, но и любой реальной области насыщения слов длины $> L$. Это означает, что гипотеза 2 верна при $L \leqslant 26$ (об этом говорится в следствии). А при $L' > 26$ для проверки данной гипотезы достаточно рассмотреть точки $D^*_{L'}$, не пренадлежащие D_{26}, т. е. точки $D^*_{L'}$, лежащие между прямыми $y = 27 - n$ (границей области D^*_{26}) и $y = L' + 1 - n$.

А теперь рассмотрим схему доказательства 3.

Этап первый: доказывается, что $D^*_1 \subseteq D_{L'}$ для всех $L' \geqslant 1$, то есть, что теорема верна для $L = 1$.

Этап второй: доказывается, что $\Delta D^*_2 \subseteq D_{L'}$ для всех $L' \geqslant 2$,

Напомним, что $D^*_L = \Delta D^*_L \bigcup D^*_{L-1}$. Поэтому, на основании первого и второго этапов делается вывод, что теорема верна для $L = 2$, т. е. $D^*_2 \subseteq D_{L'}$ для всех $L' \geqslant 2$.

Этап третий: доказывается, что $\Delta D^*_3 \subseteq D_{L'}$ для всех $L' \geqslant 3$. На основании последних двух этапов делается вывод, что теорема верна для $L = 3$, т. е. $D^*_3 \subseteq D_{L'}$ для всех $L' \geqslant 3$.

Аналогично доказывается, что теорема верна для $L = 4, 5, \ldots, 26$.

Доказательство заключается в том, что, начиная с 2 (для $L = 1$ доказательство несколько отличается), для каждого значения L, для каждой точки из рассматриваемого на этом этапе множества приведено слово длины L, которое насыщается в этой точке, удовлетворяющее условию 3. Из этого слова методом, описанным в 3, можно получить слово любой длины $> L$, имеющее ту же точку насыщения. Таким образом, рассматриваемая точка является точкой насыщения слов длины $\geqslant L$. Вид слова, выбираемого в качестве примера определяется по табл. 1.

Список литературы

[1] М. Т. Тихомирова, Функция сложности и области насыщения 01-слов // Матем. сист. – Красноярск: КрасГАУ. – 2005. – №4. (в печати).

[2] Н. В. Благодатская, Функции сложности и области насыщения 01-слов, Красноярск: КГПУ, 2004.

Красноярский государственный аграрный университет
E-mail: tihomirova_mv@mail.ru