ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Октябрь—декабрь 2004. Серия 1. Том 11, № 4, 20–35

УДК 519.114

СТАТИСТИЧЕСКИЕ ЗАКОНОМЕРНОСТИ ВЗАИМОДЕЙСТВИЯ ПЕРИОДОВ ЧАСТИЧНЫХ СЛОВ

Ю. В. Гамзова

Частичное слово длины n над алфавитом A есть частичная функция $W : \{1, \ldots, n\} \to A$. Частичное слово рассматривают как обычное слово над алфавитом $A_\diamond = A \cup \{\diamond\}$, полагая $W(i) = \diamond$ для i таких, что $W(i)$ не определена. Символ $\diamond$ называется *джокером*.
Частичное слово W имеет *период* p, если $W(i) = W(j)$ для всех $W(i), W(j) \in A$, $i \equiv j \pmod p$. *Свойство взаимодействия периодов* для периодических слов заключается в следующем: слово с периодами p и q имеет также период НОД(p, q). Выполнение этого свойства для обычных слов зависит только от длины слова, а для частичных слов — от длины слова, а также от числа и расположения джокеров в слове. В данной статье исследуется случай, когда наличие свойства взаимодействия периодов не обусловлено достаточно большой (по сравнению с числом джокеров) длиной, т. е. это свойство может присутствовать или отсутствовать в зависимости от расположения джокеров в слове. Разработан полиномиальный (с фиксированным параметром) алгоритм для определения вероятности выполнения свойства взаимодействия периодов для частичных слов данной длины с данным числом джокеров и приведены результаты полученных экспериментов.

Введение

Частичное слово представляет собой комбинаторный объект, получаемый естественным обобщением понятия обычного конечного слова. На слово длины n можно смотреть как на функцию из множества $\{1, \ldots, n\}$ в алфавит. Тогда частичное слово длины n над алфавитом A может быть определено (см. [2]) как *частичная* функция

$$W : \{1, \ldots, n\} \to A.$$

Комбинаторика частичных слов находится в начальной стадии своего развития и большинство постановок задач являются прямыми обобщениями задач комбинаторики «обычных» слов.

В данной статье продолжается исследование фундаментального свойства периодических слов — свойства взаимодействия периодов, начатое в [1, 2].

Дадим необходимые определения для обычных слов. Под *длиной* слова W понимается число символов в слове W (обозначается через $|W|$). Число символов $a \in A$ в слове W обозначается через $|W|_a$. Порядковый номер символа в слове называется его *позицией*. Натуральное число p, не превосходящее $|W|$, называется *периодом* слова W, если

$$W(i) = W(i+p), \qquad i = 1, \dots, |W| - p.$$

В частности, любое слово имеет тривиальный период, равный его длине. В дальнейшем мы будем рассматривать слова, имеющие по меньшей мере два различных нетривиальных периода. Для периодических слов одним из основных свойств является свойство взаимодействия периодов: достаточно длинное слово с периодами p и q имеет также и «производный» период НОД(p, q). Точная формулировка приводится в теореме Файна–Вильфа, которую мы приведём в удобных для нас обозначениях.

Теорема 1 [4]. *Пусть p и q — натуральные числа. Тогда каждое слово длины не менее $p + q-$НОД(p, q) с периодами p и q имеет период* НОД(p, q). *Эта оценка является точной.*

Перейдем к частичным словам. Частичное слово над алфавитом A будем представлять как обычное слово над расширенным алфавитом $A_\diamond = A \cup \{\diamond\}$, подставляя символ $\diamond$ в неопределенные позиции. Символ $\diamond$ будем называть *джокером*.

Частичное слово W имеет *период p*, если для всех позиций, не занятых джокерами, выполняется условие

$$i \equiv j \, (\mathrm{mod} p) \Rightarrow W(i) = W(j).$$

Например, слово $W = ab\diamond a\diamond cabc$ имеет период 3. Заметим, что частичное слово имеет период p тогда и только тогда, когда в нем все джокеры можно заменить буквами так, что полученное обычное слово будет иметь период p.

Как отмечалось в разных работах (см., например, [2, 3]), утверждения о взаимодействии произвольных периодов p и q тривиально следуют из аналогичных утверждений для взаимно простых периодов ввиду следующего наблюдения.

Наблюдение. В случае, когда НОД$(p, q) = d > 1$, частичное слово U можно заменить набором, состоящим из d (частичных) слов $U_1, \dots, U_d$, где $U_i = U(i)U(d+i)U(2d+i)\ldots$. Каждое из этих слов будет иметь взаимно простые периоды p/d, q/d. При этом слово U будет иметь период d тогда и только тогда, когда все U_i имеют период 1.

Слово, имеющее период p, не может содержать более p различных букв, т. е. период накладывает ограничение на максимальное возможное число букв в слове. В случае взаимно простых периодов свойство взаимодействия периодов можно переформулировать так: все буквы в достаточно длинном слове с периодами p и q должны быть одинаковы. Таким образом, свойство взаимодействия периодов заключается в том, что два периода, взаимодействуя, накладывают на слово более сильные ограничения, чем каждый из них по отдельности. Свойство взаимодействия периодов можно рассмотреть в обобщенном варианте: возможное число различных букв в достаточно длинном слове с периодами p и q ограничено некоторой константой, которая меньше длины минимального периода.

Выполнение свойства взаимодействия периодов зависит только от некоторых параметров слова (для обычных слов — от длины слова, а для частичных слов — от длины слова, числа и расположения джокеров в нем). Поэтому при изучении свойства взаимодействия периодов вместо частичных слов будут рассматриваться расстановки. *Расстановка* A — это слово над $\{0, 1\}$, полученное из частичного слова W с помощью следующего гомоморфизма:

$$A(i) = \begin{cases} 1 & \text{при } i \in D(W), \\ 0 & \text{при } i \notin D(W). \end{cases}$$

Максимум из числа различных букв в соответствующих данной расстановке частичных словах с периодами p и q будем называть *размерностью* расстановки.

Пусть взаимно простые периоды p и q фиксированы. Будем обозначать через (r, L, k) множество всех расстановок длины L, содержащих k джокеров и имеющих размерность r. Если какой-то параметр заменен знаком $\forall$, то это означает, что рассматривается объединение таких множеств по всем возможным значениям параметра. Для удобства мы будем писать «расстановка вида (r, L, k)» или «расстановка (r, L, k)» вместо «расстановка из множества (r, L, k)».

В работе [1] приведена оценка минимальной длины $L(k, p, q)$, при которой свойство взаимодействия периодов выполняется в классической формулировке, т. е. все расстановки, содержащие k джокеров, имеют размерность 1.

Теорема 2 [1]. *Пусть* $p > q \geqslant 3$, $k \geqslant \lfloor 2p/3 \rfloor - 1$ *при* $q{=}3$ *и* $k \geqslant \lfloor 3p/q \rfloor + 3$ *при* $q{\geqslant}4$, $L(k,p,q)$ *— длина взаимодействия периодов* p *и* q *при наличии* k *джокеров. Тогда*

$$L(k,p,q) = \frac{pqk}{p+q-2} + \Delta(k,p,q),$$

где функция $\Delta(k,p,q)$ *является периодической по* k *с периодом* $p+q-2$ *и*

$$\frac{pq}{p+q-2} \leqslant \Delta(k,p,q) < 4(q-1).$$

Таким образом, поведение длины взаимодействия периодов как функции от числа джокеров известно. В данной статье исследуются расстановки, длина которых меньше длины взаимодействия. В этом случае расстановки одинаковой длины с одинаковым числом джокеров могут иметь различную размерность, и интересным представляется вопрос о доле расстановок определенной размерности.

Обозначим через $P(r,L,k)$ долю расстановок длины L, содержащих k джокеров и имеющих размерность r, т. е. отношение мощности множества (r,L,k) к мощности множества $(\forall,L,k)$. Разработан эффективный алгоритм для определения $P(r,L,k)$. Время работы «наивного» алгоритма экспоненциально зависит от длины слова; время работы предлагаемого алгоритма полиномиально зависит от L и p и экспоненциально — от параметра q. Данная статья посвящена изложению этого алгоритма и результатов, полученных с его помощью.

В следующих двух разделах излагается идея алгоритма. В четвертом разделе рассматривается практическая реализация алгоритма и оценка его сложности. В последнем разделе приводятся результаты и гипотезы, связанные с полученными результатами.

2. Идея алгоритма

В данной статье предлагается алгоритм для вычисления $P(r,L,k)$, имеющий полиномиальную сложность по L и p и экспоненциальную по q. Алгоритм основан на разбиении расстановок (r,L,k) на классы и вычислении их числа в каждом классе.

Пусть $l = \lceil |A|/pq \rceil$. Дополним расстановку A нулями до длины lpq. Тогда ее можно представить как конкатенацию расстановок $A_i, i = 1, \dots l$, где $|A_i| = pq$. Применим к словам $A_i, i = 1, \dots l$, последовательно логическое ИЛИ. Полученную в результате расстановку длины pq назовем *характеристической* расстановкой.

Пример 1. Рассмотрим частичное слово с периодами 4 и 5 и соответствующую ему расстановку:

$$aa\diamond\diamond\diamond\diamond\diamond ba\diamond\diamond\diamond\diamond\diamond\diamond\diamond\diamond\diamond a\diamond\diamond\diamond\diamond\diamond\diamond\diamond\diamond baa\diamond\diamond$$
$$11000001100000000010000000011100$$

Характеристическая расстановка получается из исходной расстановки следующим образом.

1. Исходная расстановка дополняется нулями до длины, кратной pq, и получившееся слово представляется как конкатенация слов A_1 и A_2, $|A_1| = |A_2| = pq$:

$$\underbrace{11000001100000000010}_{A_1}\underbrace{000000011100\ \ 00000000}_{A_2}.$$

2. К словам A_1 и A_2 последовательно применяется логическое ИЛИ, в результате чего получается характеристическая расстановка

$$\begin{array}{rl} & A_1 = 11000001100000000010 \\ \vee & A_2 = \underline{00000001110000000000} \\ & 11000001110000000010 \end{array}$$ — характеристическая расстановка.

Утверждение 1. *Размерность расстановки совпадает с размерностью ее характеристической расстановки.*

Доказательство. Заметим, что для каждого периода p произвольного слова U и для любого i, $1 \leqslant i \leqslant p$, все буквы, находящиеся в позициях $i, i+p, i+2p, \ldots$, одинаковы независимо от расположения джокеров в этих позициях. Указанное множество позиций будем называть *p-классом*; оно содержится в i-м классе вычетов по модулю p.

Мы рассматриваем слова с двумя различными периодами p, q у слова U и соответственно p- и q-классы. Заметим, что если две позиции из разных q-классов, занятые буквами, принадлежат одному p-классу (т. е. «связаны» при помощи периода p), то в этих q-классах содержится одна и та же буква. Если же любая пара позиций из данных q-классов, принадлежащая одному p-классу, содержит хотя бы один джокер, то эти q-классы могут содержать различные буквы.

Таким образом, размерность расстановки зависит от того, какие q-классы связаны между собой при помощи периода p. Заметим, что в характеристической расстановке единица присутствует в позиции, принадлежащей данным q- и p-классам тогда и только тогда, когда в исходной расстановке единица присутствует хотя бы в одной из позиций, принадлежащих тем же q- и p-классам. Поэтому в характеристической расстановке q- и p-классы связаны между собой так же, как в исходной, и, значит, размерности исходной и характеристической расстановок совпадают. Утверждение 1 доказано.

Пример 2. Продолжим рассмотрение примера 1. В соответствующих характеристической расстановке 1100000111000000001 0 словах множество позиций, в которых находятся буквы, состоит из 6 элементов: $D = \{1, 2, 8, 9, 10, 19\}$. Будем нумеровать 4-классы (5-классы) остатком от деления на 4 (соответственно на 5) номеров позиций, принадлежащих этому классу. Первая и девятая позиции принадлежат первому 4-классу, девятая и девятнадцатая — первому 5-классу. Значит, в этих позициях должна стоять одна и та же буква. Вторая и десятая позиции принадлежат второму 4-классу. Больше связей с помощью периодов между позициями в D нет. Значит, соответствующие характеристической расстановке частичные слова могут содержать не более 3 различных букв (например, $ab\diamond\diamond\diamond\diamond\diamond cab\diamond\diamond\diamond\diamond\diamond\diamond\diamond\diamond a\diamond$). Поэтому размерность характеристической (и, значит, исходной) расстановки равна 3.

Каждой расстановке поставим в соответствие ее характеристическую расстановку и разобьем расстановки на классы. Тогда для определения $P(r, L, k)$ необходимо найти все характеристические расстановки размерности r, а затем для каждой такой характеристической расстановки требуется определить число соответствующих ей расстановок (r, L, k). Характеристические расстановки можно находить, например, перебором (вычислительная сложность такого перебора зависит только от p и q и не зависит от длины L, которая может быть сколь угодно большой по сравнению с периодами). Тем не менее на практике и такой перебор достаточно затруднителен. Поэтому далее будет рассмотрен эффективный способ подсчета характеристических расстановок заданной размерности.

3. Число расстановок с заданной характеристической расстановкой

Сначала рассмотрим задачу определения мощности класса расстановок, которым соответствует данная характеристическая расстановка. Из ее решения будет следовать, что можно не перечислять все характеристические расстановки, а разбить их на классы так, что характеристическим расстановкам из одного класса соответствует одно и то же число расстановок, а затем подсчитать число характеристических расстановок в каждом классе. Возможны два варианта этой задачи:

1. Найти число расстановок $(\forall, L, \forall)$ с данной характеристической расстановкой.

2. Найти число расстановок $(\forall, L, k)$ с данной характеристической расстановкой.

Начнем с решения первого варианта задачи. Известно, что в позициях исходной расстановки, соответствующих позициям нулей в характеристической расстановке, должны находиться нули, в остальных позициях могут находиться как нули, так и единицы, но каждой единице в характеристической расстановке должна соответствовать хотя бы одна единица в исходной расстановке.

Рассмотрим случай $L = lpq$. В этом случае каждой единице в характеристической расстановке H соответствуют l позиций в исходной расстановке. В каждой из этих позиций может находиться 0 или 1, что дает 2^l комбинаций. Однако во всех позициях нули находиться не могут. Поэтому каждой единице характеристической расстановки соответствуют $2^l - 1$ комбинаций. Пусть $h_1 = |H|_1$ и $h_0 = |H|_0$. Таким образом, число расстановок длины $L = lpq$, которым соответствует данная характеристическая расстановка H, равно $(2^l - 1)^{h_1}$.

В случае, когда $L = lpq + b, 0 < b < pq$, обозначим через H' префикс слова H длины b. Тогда число расстановок длины $L = lpq + b$, которым соответствует данная характеристическая расстановка H, равно

$$(2^{l+1} - 1)^{|H'|_1} + (2^l - 1)^{h_1 - |H'|_1}.$$

Таким образом, решение первой задачи можно получить с помощью указанных выше формул за время $O(pq)$, требуемое для вычисления h_0 и h_1.

Второй вариант задачи существенно сложнее, так как рассматриваются не все расстановки, которым соответствует данная характеристическая расстановка, а только те, в которых содержится ровно k нулей.

Пусть $L = lpq$. В позициях, соответствующих нулям в характеристической расстановке H, в исходной расстановке также должны находиться нули. Число таких позиций в исходной расстановке равно lh_0. Оставшиеся $k' = k - lh_0$ нулей нужно расставить в позициях, соответствующих единицам в H. Для этого рассмотрим все такие (неупорядоченные) представления числа k' в виде суммы h_1 неотрицательных целых слагаемых, что каждое слагаемое не превосходит $l - 1$. Тогда i-е слагаемое в представлении соответствует i-й по счету единице в характеристической расстановке и равно числу нулей, которые необходимо поставить в исходной расстановке в позициях, соответствующих позиции этой единицы. Так как число таких позиций равно l , то число возможных расстановок нулей, соответствующих слагаемому, равному z, равно $\binom{l}{z}$.

Через $d(k', h_1)$ обозначим число расстановок $(\forall, L, k)$, соответствующих данной характеристической расстановке H. При фиксированном l эта величина зависит только от k' и числа единиц в H (и не зависит от периодов p, q).

Через $K_{i,j}$ обозначим множество всех представлений числа i в виде суммы j неотрицательных слагаемых, каждое из которых не превосходит $l-1$, а через z_{nm} — m-е слагаемое из n-го представления. Тогда

$$d(k', h_1) = \sum_{n=1}^{|K_{k',h_1}|} \prod_{m=1}^{h_1} \binom{l}{z_{nm}}.$$

В случае, когда $L = lpq + b, 0 < b < pq$, необходимо учитывать, что единицам, принадлежащим первым b позициям характеристической расстановки, соответствуют $l+1$ позиций исходной расстановки. Обозначим через s число единиц, находящихся в первых b позициях характеристической расстановки. Тогда для заданной характеристической расстановки H число соответствующих ей расстановок $(\forall, L, k)$ равно

$$d(k', h_1) = \sum_{n} \prod_{m=1}^{s} \binom{l+1}{z_{nm}} \prod_{m=s+1}^{h_1} \binom{l}{z_{nm}},$$

где суммирование производится по всем представлениям числа k' в виде суммы h_1 слагаемых, из которых первые s не превосходят l, а остальные не превосходят $l-1$.

В следующем разделе будет представлен алгоритм вычисления $d(i, j)$, время работы которого полиномиально зависит от p и экспоненциально от q.

Итак, в случае $L = lpq$ для определения числа расстановок (r, L, k) для всех $x = 1, \ldots, pq$ достаточно знать число расстановок (r, pq, x). В случае $L = lpq + b, 0 < b < pq$, достаточно знать для всех $x = 1, \ldots, pq$ и для всех $y = 1, \ldots, x$ число расстановок (r, pq, x) таких, что в первых b позициях находятся y единиц.

4. Построение вспомогательных таблиц

При вычислении $P(r, L, k)$ для различных r и k можно построить один раз таблицу B размера $q \times pq$, где $b(i, j)$ равно числу характеристических расстановок с размерностью i и числом единиц j, и таблицу D размера $k' \times pq$, элементы $d(i, j)$ которой определены выше. Тогда

$$P(r, L, k) = \frac{\sum\limits_{i=0}^{q} b(r, i) d(k - l(pq - i), i)}{\binom{L}{k}}, \qquad (*)$$

т. е. при наличии этих таблиц временна́я сложность вычисления $P(r, L, k)$ линейно зависит от q. В этом параграфе рассмотрены алгоритмы построения таблиц B и D и временна́я сложность этих алгоритмов.

В дальнейшем для упрощения изложения будем рассматривать случай $L = lpq$. Характеристическую расстановку H будем записывать не в виде слова длины pq, а в виде таблицы размера $p \times q$ следующим образом: каждый элемент расстановки $H(n)$ записывается в клетке (i, j), где $i = n \pmod p), j = n \pmod q)$. Полученную таблицу назовем *бланком*. Единицы, находящиеся в одной строке (столбце) бланка, соответствуют позициям, равным по модулю p (по модулю q). Поскольку p, q — периоды, то все буквы в позициях, принадлежащих одной строке или одному столбцу, должны быть одинаковыми.

Перестановками строк и столбцов приведем бланк к клеточно-диагональному виду с максимальным возможным числом клеток (эти клетки будем называть *блоками*). Ограничение на перестановки таково: порядок следования строк (столбцов) внутри блоков должен быть таким же, как в исходном бланке. Далее блоки переставим по убыванию высоты, а одинаковые по высоте — по убыванию ширины. Полученную таблицу будем называть *фактор-бланком*.

Пример 3. Вернемся к рассмотрению примера 1. Представим характеристическую расстановку 11000001110000000010 в виде бланка. Первый символ расстановки будет записан в клетке (1,1), второй — в клетке (2,2), третий — в клетке (3,3), четвертый — в клетке (4,0), пятый — в клетке (0,1) и т. д. Результат приведен на рис. 1a. При перестановке строк и столбцов так, как указано на рис.1b, получается фактор-бланк.

	0	1	2	3
0	0	0	1	0
1	0	1	0	0
2	0	0	1	0
3	1	0	0	0
4	0	1	0	1

a)

	1	3	2	0
1	1	0	0	0
4	1	1	0	0
0	0	0	1	0
2	0	0	1	0
3	0	0	0	1

b)

Рис. 1. Представление расстановки 11000001110000000010 в виде бланка a) и соответствующий фактор-бланк b)

Утверждение 2. *Число блоков в фактор-бланке совпадает с размерностью исходной расстановки.*

Доказательство. Сначала покажем, что две единицы принадлежат одному блоку тогда и только тогда, когда их можно соединить такой ломаной, что она начинается в клетке первой единицы, заканчивается в клетке второй единицы, и в каждой клетке излома стоит единица. В противном случае перестановкой строк и столбцов этот блок можно было бы привести к клеточно-диагональному виду не менее чем с двумя клетками, что противоречит максимальности числа блоков в фактор-бланке.

Если две единицы фактор-бланка можно соединить ломаной, построенной указанным выше способом, то это означает, что в соответствующем данной расстановке частичном слове буквы, находящиеся в позициях этих единиц, равны, так как эта ломаная соединяет между собой позиции, принадлежащие одной и той же строке или столбцу, т. е. равные по модулю p или q. В противном случае в соответствующих позициях могут находиться разные буквы. Это означает, что в соответствующем данной расстановке частичном слове в позициях из одного блока должны находиться одинаковые буквы, в позициях из разных блоков могут находиться разные буквы. Поэтому число блоков в фактор-бланке совпадает с размерностью исходной расстановки. Утверждение 2 доказано.

Приступим к построению таблицы B. Все бланки размера $i \times j$ с n единицами (число таких бланков равно $\binom{ij}{n}$) можно разделить на:

(1) бланки, фактор-бланк которых содержит более одного блока (число таких бланков обозначим через $s_m(i,j,n)$, где m — число блоков в фактор-бланке);

(2) блоки размера $i' \times j'$ с n единицами, где $i' \leqslant i$, $j' \leqslant j$, $i'j' < ij$, дополненные пустыми строками или столбцами до размера $i \times j$ (число таких бланков обозначим через $t_{i',j'}(i,j,n)$);

(3) блоки размера $i \times j$ с n единицами (число таких бланков обозначим через $e(i,j,n)$).

Каждый элемент таблицы B вычисляется независимо от остальных по данным вспомогательных таблиц E, S_m $(m = 2,\ldots,q)$ и $T_{i',j'}$ $(i' = 1,\ldots,q,\ j' = 1,\ldots,p)$. Размер каждой из этих таблиц равен $p \times q \times pq$.

Для построения таблицы E используем метод динамического программирования. Поскольку $e(i,j,n) = e(j,i,n)$, достаточно найти только те элементы таблицы E, которые находятся над и на главной диагонале. Начнем с минимального блока размера 1×1, содержащего одну единицу. При построении таблицы E переход от элемента к элементу осуществляется увеличением на единицу первого справа индекса, для которого это возможно.

Бланки вида (1) получаются из фактор-бланка, содержащего несколько блоков, с помощью перестановки строк и столбцов с сохранением порядка строк и столбцов внутри блоков, т. е. для каждой строки и столбца бланка достаточно указать номер соответствующего блока. При такой перестановке строк и столбцов одинаковые бланки при разных перестановках могут получиться только в случае наличия одинаковых блоков. Для оптимизации будем использовать дополнительные таблицы F_m, где элемент $f_m(i,j,n)$ равен числу бланков длины ij с n единицами вида (1) без пустых строк и столбцов, фактор-бланк которых содержит m блоков. Пусть фактор-бланк состоит из блоков $B_1,\dots,B_m, m>1$, каждый блок имеет размер $x_1\times y_1,\dots,x_m\times y_m$, $\sum\limits_{g=1}^{m} x_g n_g = i$, $\sum\limits_{g=1}^{m} y_g n_g = j$, и встречается $n_1,\dots n_m$ раз соответственно. Тогда число бланков, соответствующих этому фактор-бланку, равно $\binom{i}{x_1n_1,\dots,x_mn_m}\binom{j}{y_1n_1,\dots,y_mn_m}$. Таким образом, для определения $f_m(i,j,n), m>1$, необходимо рассмотреть все возможные комбинации m блоков, сумма размеров которых равна $i\times j$, а сумма единиц равна n. Для каждой из этих комбинаций блоков можно определить число соответствующих ему бланков без пустых строк и столбцов по указанной выше формуле. Все бланки длины ij с n единицами вида (1) получаются из бланков длины $i'j'$, $i'j'\leqslant ij$ с n единицами с помощью добавления пустых строк и/или столбцов, поэтому

$$s_m(i,j,n)=\sum_{i'=1}^{i}\sum_{j'=1}^{j} f_m(i',j',n)\binom{i}{i'}\binom{j}{j'}, \quad m=2,\dots,q.$$

При построении таблицы F_m для $m=2,\dots,q$ происходит перебор всех возможных комбинаций параметров m блоков (т. е. высоты, ширины и числа единиц). Каждая такая комбинация рассматривается не более одного раза. Число блоков, имеющих заданные параметры, определяется по уже построенной части таблицы E (элемент $e(x,y,n)$ равен числу блоков размера $x\times y$ с n единицами). Поэтому число необходимых операций при этом ограничено произведением чисел разбиений (т. е. представлений в виде суммы неупорядоченных натуральных слагаемых) чисел p, q и pq, причем эти разбиения могут содержать не более q частей. Таким образом, временна́я сложность построения таблиц $F_m, m=2,\dots,q$ не превосходит $O((pq)^{2q})$. Величина $s_m(i,j,n)$ вычисляется по данным таблицы F_m за полиномиальное от p,q время.

Бланки вида (2) получаются добавлением пустых строк и столбцов к блокам с n единицами, размер которых меньше $i\times j$. Число этих блоков известно, так как в таблице E заполнены все элементы вплоть до элемента $e(i,j,n)$. Из блока размера $x\times y$ с n единицами получается $\binom{i}{x}\binom{j}{y}$

бланков размера $i \times j$ вида (2). Таким образом,

$$t_{x,y}(i,j,n) = e(x,y,n)\binom{i}{x}\binom{j}{y}.$$

Нахождение числа бланков вида (2) также производится за полиномиальное от p, q время.

Итак, для определения очередного элемента таблицы E (т. е. числа бланков вида (3)) необходимо найти разность $\binom{ij}{n}$ и числа бланков вида (1) и (2):

$$e(i,j,n) = \binom{ij}{n} - \sum_{m=2}^{q} s_m(i,j,n) - \sum_{i'=1}^{i-1}\sum_{j'=1}^{j-1} t_{i',j'}(i,j,n) - \sum_{i'=1}^{i-1} t_{i',j}(i,j,n) - \sum_{j'=1}^{j-1} t_{i,j'}(i,j,n).$$

Элементы таблицы B получаются при вычислении последних pq элементов таблицы E (для всех $n = 1, \ldots, pq$ элементы $b(m,n), m = 1, \ldots, q$ получаются при вычислении $e(p,q,n)$) следующим образом:

$$b(1,n) = e(p,q,n) + \sum_{i'=1}^{p}\sum_{j'=1}^{q} t_{i',j'}(p,q,n) - t_{p,q}(p,q,n),$$
$$b(m,n) = s_m(p,q,n), \quad m = 2, \ldots, q.$$

Таким образом, элемент $b(1,n)$ равен сумме числа бланков вида (1) и (2). Элемент $b(m,n)$, $m = 2, \ldots, q$, равен числу бланков вида (3), в которых фактор-бланк состоит ровно из m блоков. Временна́я сложность выполнения указанных выше операций для построения таблицы E полиномиально зависит от p и экспоненциально от q.

Для построения таблицы D также используем метод динамического программирования. Напомним, что элемент $d(i,j)$ равен числу расстановок $(\forall, L, l(pq-i)+j)$, характеристической расстановкой которых является данная расстановка H с i единицами, причем эта величина не зависит от выбора H. Заметим, что $d(0,j) = 1, j = 1, \ldots, pq$, поскольку имеется ровно одно представление 0 в виде любого числа неотрицательных слагаемых, $\binom{l}{0} = 1$, $l = 1, \ldots, pq$. Положим $d(0,0) = 1$ (этот элемент соответствует «пустому месту»), $d(i,0) = 0, i > 0$. Обозначим через $K_{i,j}$ множество всех представлений числа i в виде суммы j неотрицательных слагаемых, каждое из которых не превосходит $l-1$, и пусть z_{nm} равно m-у слагаемому из n-го представления. По определению

$$d(i,j)=\sum_{n=1}^{|K_{ij}|}\prod_{m=1}^{j}\binom{l}{z_{nm}},$$

где суммирование производится по всем представлениям из $K_{i,j}$. Каждому представлению из $K_{i,j}$, в котором последнее слагаемое равно x, можно поставить во взаимно однозначное соответствие представление из $K_{i-x,j-1}$ (первое представление получается из второго с помощью «дописывания» справа слагаемого x). Для первого представления величина $\prod_{m=1}^{j}\binom{l}{z_{nm}}$ равна произведению той же величины из второго представления на $\binom{l}{x}$. Представления (i,j) разобьем на l классов в соответствии с величиной последнего слагаемого представления (от 0 до $l-1$). Тогда

$$d(i,j)=\sum_{m=\min\{0,i-l+1\}}^{i}d(m,j-1)\binom{l}{(i-m)}.$$

Для определения каждого элемента таблицы D необходимо просмотреть не более l элементов. Высота таблицы равна k, ширина равна pq. Таким образом, временна́я сложность построения таблицы D равна $O(kpql)=O(L^2)$.

Итак, построив таблицы B и D, можно вычислить $P(r,L,k)$ за время $O(q)$ по формуле $(*)$. Заметим, с одними и теми же таблицами B, D можно вычислить $P(r,L,k')$ при всех $k'\leqslant k$. Более того, при вычислении $P(r,L',k')$ при $L'\neq L$ требуется перестраивать только таблицу D.

5. Полученные результаты

Программа, реализующая изложенный выше алгоритм, использовалась для получения эмпирических данных о поведении $P(r,L,k)$ при больших (по сравнению с периодами) длинах (эти данные невозможно было получить с помощью «наивного» алгоритма, так как перебор при этом был слишком велик). В частности, исследовалась зависимость доли неунарных расстановок (т. е. расстановок, размерность которых больше единицы) длины L с k джокерами от доли джокеров в слове (т. е. от величины k/L). Все графики данной зависимости очень похожи (см. примеры на рис. 2, 3). Вертикальная линия на обоих графиках соответствует $\dfrac{k}{L}=\dfrac{p+q-2}{pq}$ — оценке длины взаимодействия из теоремы 2.

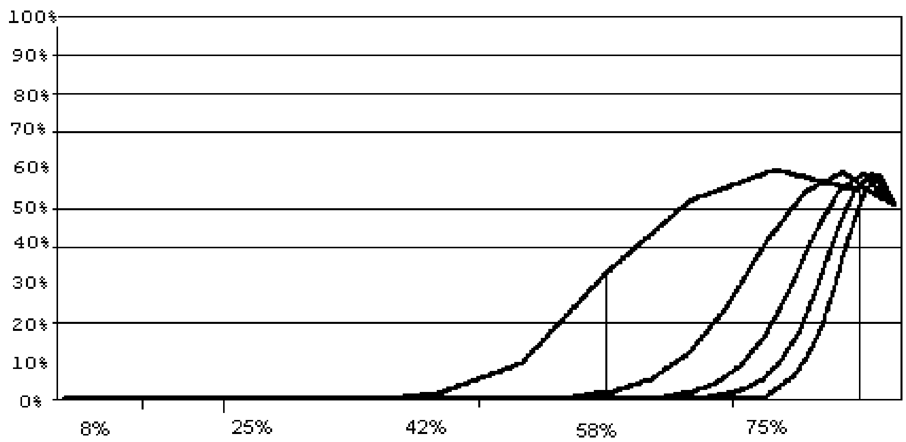

Рис. 2. Зависимость доли неунарных расстановок от доли джокеров в случае $q = 3, p = 4, L = 12, 24, 36, 48, 60$ (слева направо)

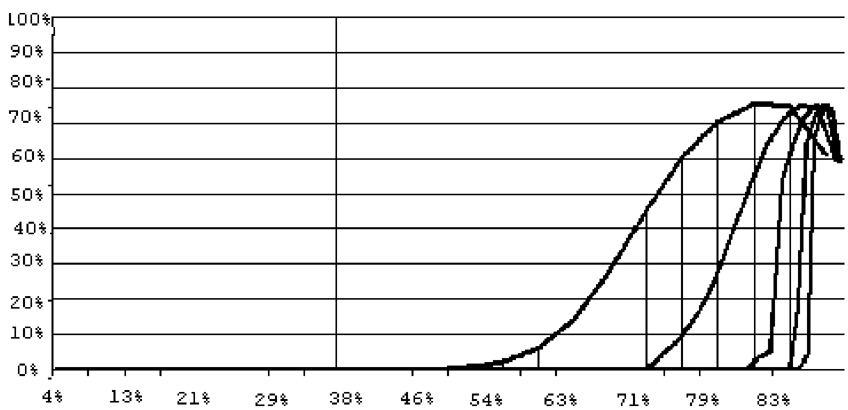

Рис. 3. Зависимость доли неунарных расстановок от доли джокеров в случае $q = 3, p = 8, L = 24, 48, 72, 96, 120$ (слева направо)

При исследовании полученных графиков были выявлены закономерности, в том числе достаточно неожиданные, и сформулированы две гипотезы. Их изложением мы и заканчиваем данную работу.

Закономерность 1. Доля неунарных расстановок близка к нулю на большом отрезке графика «справа» от длины взаимодействия.

Закономерность 2. Доля джокеров, при которой начинается заметный рост доли неунарных расстановок, при увеличении длины также увеличивается. Таким образом, если зафиксировать ε, $0 < \varepsilon < 1$, то при увеличении длины число джокеров, при котором доля неунарных расстановок начинает превышать ε, увеличивается не пропорционально увеличению длины, а на бо́льшую величину.

Можно сформулировать следующую гипотезу.

Гипотеза 1. Пусть $\varepsilon \in (0,1)$ — фиксированное число, $L_1 < L_2$. Через k_1 и k_2 обозначим число джокеров такое, что $P(r, L_1, k_1) > \varepsilon$, $P(r, L_1, k_1 - 1) \leqslant \varepsilon$ и $P(r, L_2, k_2) > \varepsilon$, $P(r, L_2, k_2 - 1) \leqslant \varepsilon$.

Тогда $\dfrac{k_1}{L_1} < \dfrac{k_2}{L_2}$.

Закономерность 3. Доля джокеров, при которой достигается максимум доли неунарных расстановок, при увеличении длины увеличивается. Более того, в рассмотренных случаях максимум доли неунарных расстановок достигался при $L - k = \mathrm{const} \leqslant q$ (т. е. одном и том же числе единиц в расстановке), причем само максимальное значение не зависит от длины.

Гипотеза 2. Пусть $k^*(L)$ — число джокеров, при котором достигается максимум доли неунарных расстановок длины L. Тогда существуют числа $0 < \beta(q,p) \leqslant q$, $0 < \gamma(q,p) \leqslant 1)$, такие, что для любой длины L

$$k^*(L) = L - \beta(q,p), \quad \sum_{r=2}^{q} P(r, L, k^*(L)) = \gamma(q,p).$$

Автор выражает благодарность А. М. Шуру за всестороннюю поддержку.

ЛИТЕРАТУРА

1. **Шур А. М., Гамзова Ю. В.** Частичные слова и свойство взаимодействия периодов // Изв. РАН. Серия матем. 2004. Т. 68, № 2. С. 199–222.
2. **Berstel J., Boasson L.** Partial words and a theorem of Fine and Wilf // Theoret. Comput. Sci. 1999. V. 218, N 1. P. 135–141.

3. **Choffrut C., Karhumäki J.** Combinatorics on words // Handbook of formal languages. V. 1. Berlin: Springer, 1997. P. 329–438.

4. **Fine N. J., Wilf H. S.** Uniqueness theorem for periodic functions // Proc. Amer. Math. Soc. 1965. V. 16, N 1. P. 109–114.

Адрес автора:
Уральский гос. университет,
пр. Ленина, 51,
620083 г. Екатеринбург, Россия.
E-mail: snarl@r66.ru

Статья поступила
18 августа 2004 г.