ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Апрель—июнь 2000. Серия 1. Том 7, № 2, 65–70

УДК 519.115.1

ОЦЕНКИ ДЛИНЫ УНИВЕРСАЛЬНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ ДЛЯ ПЕРЕСТАНОВОК*)

В. Ню, Д. Фон-Дер-Флаасс

Получены оценки длины $L(n)$ минимальной универсальной последовательности для множества всех перестановок из n символов. Показано, что

$$n! + (n-1)! + \frac{n-1}{2n-3}(n-2)! + n - 3 \leqslant L(n) \leqslant n! + (n-1)! + \ldots + 1!.$$

Приведены две конструкции одной и той же универсальной последовательности длины $n! + (n-1)! + \ldots + 1!$.

Последовательность W называется *универсальной* для множества слов S, если каждое слово из S встречается в W как подслово, т. е. $W = W_1 s W_2$ для любого $s \in S$. Универсальную последовательность наименьшей длины будем называть *минимальной*. Известно [2], что для произвольного множества S задача построения минимальной универсальной последовательности является NP-полной.

Пусть $P(n)$ — множество всех перестановок символов $1, 2, \ldots, n$, а $L(n)$ — длина минимальной универсальной последовательности для $P(n)$. В [3] анонсирована оценка $L(n) \leqslant 2n!$ и несложно доказывается, что $L(n) \geqslant n! + (n-1)! + n - 2$. В настоящей работе получены оценки

$$n! + (n-1)! + \frac{n-1}{2n-3}(n-2)! + n - 3 \leqslant L(n) \leqslant n! + (n-1)! + \ldots + 1!.$$

Пусть $W = w_1 \ldots w_L$ — произвольная универсальная последовательность для $P(n)$. Подслово $w_i \ldots w_{i+n-1}$ называется *вхождением* перестановки Z в W, если $Z = w_i \ldots w_{i+n-1}$. Кратность вхождения перестановки Z в последовательность W обозначим через $k_W(Z)$.

Будем говорить, что перестановка $X = x_1 x_2 \ldots x_n$ *перекрывается* с перестановкой $Y = y_1 y_2 \ldots y_n$ по l символам, если $x_{n-l+1} \ldots x_n = y_1 \ldots y_l$.

*) Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (коды проектов 99-01-00531 и 99-01-00581) и INTAS (проект INTAS-OPEN-97-1001).

Число $C(X,Y) = n - 1 - l$ назовем *стоимостью* перехода от X к Y. При $l = n - 1$ перестановка Y является циклическим сдвигом перестановки X. Только в этом случае $C(X,Y) = 0$. *Стоимостью* перестановки X в универсальной последовательности W назовем величину $C_W(X) = \sum C(X,Y) + k_W(X) - 1$, где суммирование берется по всем вхождениям перестановки X в W, а при заданном вхождении X в W перестановкой Y является следующая за X перестановка. Наконец, *стоимостью* произвольного множества $V \subseteq P(n)$ в последовательности W называется величина $C_W(V) = \sum\limits_{X \in V} C(X)$. Так как из контекста обычно ясно, о какой последовательности идет речь, то индекс W, как правило, будем опускать. Непосредственно из определений вытекают следующие утверждения.

Утверждение 1. *Если* $V = V_1 \cup V_2$ *и* $V_1 \cap V_2 = \varnothing$, *то*

$$C_W(V) = C_W(V_1) + C_W(V_2).$$

Утверждение 2. $|W| = n! + C_W(P(n)) + n - 1$.

Перестановку $X_i = x_i \dots x_n x_1 \dots x_{i-1}$ будем называть i-м *циклическим сдвигом* перестановки $X = x_1 x_2 \dots x_n$, а через $X_i^{j,l}$, $1 \leqslant j \leqslant l < n$, будем обозначать i-й циклический сдвиг перестановки $X^{j,l} = x_j x_{j+1} \dots x_l x_1 \dots x_{j-1} x_{l+1} \dots x_n$, которая получена из X j-м циклическим сдвигом подслова $x_1 \dots x_l$. По определению $X^{1,l} = X$ для любого $l = 1, \dots, n$ и $X^{j,n} = X_j$.

Для удобства записи полагаем $(n)_r = n(n-1)\dots(n-r+1)$, если $1 \leqslant r \leqslant n$, и $(n)_r = 0$ в противном случае. Из определения, в частности, следует, что $(n)_n = (n)_{n-1} = n!$.

1. Построение универсальной последовательности

Пусть $X \in P(n)$ и $X = x_1 \dots x_n$. Последовательность $W_0(X) = x_1 \dots x_n x_1 \dots x_{n-1}$, содержащую перестановки $X^{1,n}, X^{2,n}, \dots, X^{n,n}$, назовем 0-*циклом*, порожденным перестановкой X.

Если для каждого i, $i < k \leqslant n-2$, построен i-цикл $W_i(X)$, то k-цикл $W_k(X)$, порожденный перестановкой X, является минимальной универсальной последовательностью, содержащей $(k-1)$-циклы $W_{k-1}(X^{1,n-k})$, $W_{k-1}(X^{2,n-k}), \dots, W_{k-1}(X^{n-k,n-k})$ в указанном порядке. Множество перестановок k-цикла будем называть k-*множеством* и обозначать через $\overline{W}_k(X)$.

Теорема 1. *Для любого* k, $0 \leqslant k \leqslant n-2$, *справедливы утверждения*:

a) $|\overline{W}_k(X)| = (n)_{k+1}$;

b) *последовательность* $W_k(X)$ *начинается перестановкой* X *и заканчивается перестановкой* $x_n x_{n-1} \dots x_{n-k} x_1 \dots x_{n-k+1}$;

c) $C(\overline{W}_k(X)) = (n-1)_k + (n-2)_{k-1} + \ldots + (n-k)_1 - k$.

Доказательство. Индукция по k. При $k = 0$ справедливость утверждений легко проверяется непосредственно.

Пусть утверждения a)–c) выполняются при всех $i < k$. Покажем, что тогда эти утверждения выполняются для $W_k(X)$.

Произвольную перестановку $y_1 \ldots y_n$ назовем *индикаторной* для X, если $y_n = x_n$. Очевидно, что 0-циклы, порожденные различными индикаторными перестановками, различны. Согласно определению $X^{j,l}$ все порождающие перестановки i-циклов, содержащиеся в $W_k(X)$, являются индикаторными. Рассмотрим порождающие перестановки двух произвольных $(k-1)$-циклов: $X^{i,n-k} = x_i \ldots x_{n-k}x_1 \ldots x_{i-1}x_{n-k+1} \ldots x_n$ и $X^{j,n-k} = x_j \ldots x_{n-k}x_1 \ldots x_{j-1}x_{n-k+1} \ldots x_n$. В расположении первых $n-k+1$ символов в этих циклах различен порядок следования соответственно в следующих парах: $x_{i-1}x_{n-k+1}$ и $x_{i-1}x_i$; $x_{j-1}x_j$ и $x_{j-1}x_{n-k+1}$. При циклических сдвигах начальных подслов длины $n-k+1$ порядок следования символов в них останется неизменным. Поэтому порождающие перестановки $(k-2)$-циклов, полученные из $X^{i,n-k}$ и $X^{j,n-k}$, различны. Аналогично доказывается различие порождающих перестановок i-циклов для $i = k-3, \ldots, 0$. Последнее, как замечено выше, означает, что все перестановки в указанных $(k-1)$-циклах различны. По предположению в каждом из них содержится $(n)_k$ перестановок. Следовательно, $W_k(X)$ содержит $(n-k) \cdot (n)_k = (n)_{k+1}$ различных перестановок. Таким образом, утверждение a) теоремы доказано.

Учитывая предположение индукции, справедливость утверждения b) легко проверяется непосредственно.

Для определения стоимости вхождения в последовательность $W_k(X)$ $(k-1)$-множества $\overline{W}_{k-1}(X^{j,n-k})$ нужно найти, учитывая предположение, только стоимость вхождения его последней перестановки $x_n \ldots x_{n-k+1}$ $x_j x_{j+1} \ldots x_{n-k}x_1 \ldots x_{j-1}$. По построению непосредственно за ней следует перестановка $X^{j+1,n-k}$, с которой она перекрывается по подслову $x_{j+1} \ldots x_{n-k}x_1 \ldots x_{j-1}$ длины $n-k-1$. Следовательно, стоимость вхождения последней перестановки равна $n-1-(n-k-1) = k$ и не зависит от j. Поскольку число таких перестановок равно $n-k-1$, то по утверждению 1 с учетом предположения имеем $C(\overline{W}_k(X)) = (n-k)C(\overline{W}_{k-1}(X)) + (n-k-1)k = (n-1)_k + (n-2)_{k-1} + \ldots + (n-k)_1 - k$. Теорема доказана.

Теорема 2. *При любом натуральном n справедливо неравенство* $L(n) \leqslant n! + (n-1)! + \ldots + 2! + 1!$.

Доказательство. Рассмотрим $(n-2)$-цикл $W_{n-2}(X)$. По теореме 1 в нем содержится $(n)_{n-1} = n!$ различных перестановок и $C(\overline{W}_{n-2}(X)) = (n-1)! + (n-2)! + \ldots + 2! - (n-2)$. Следовательно, $W_{n-2}(X)$ —

универсальная последовательность для множества $P(n)$ и по утверждению 2 она имеет длину $n! + (n-1)! + \ldots + 2! - (n-2) + (n-1) = n! + (n-1)! + \ldots + 2! + 1!$. Теорема доказана.

2. Нижняя оценка

Теорема 3. *При любом натуральном $n \geqslant 2$ справедлива оценка*

$$L(n) \geqslant n! + (n-1)! + \frac{n-1}{2n-3}(n-2)! + n - 3.$$

Доказательство. Пусть W — произвольная универсальная последовательность для множества $P(n)$. Выберем $n!$ символов, с которых начинаются вхождения в W перестановок из $P(n)$ (если какая-то перестановка встречается несколько раз, выбираем любое ее вхождение); выбранные символы покрасим в черный цвет. Последние $n-3$ символа покрасим в красный, остальные символы — в белый цвет.

Последовательность W разобьем на максимальные одноцветные отрезки; пары из черного и следующего за ним белого отрезка назовем *блоками.*

Перестановки, соответствующие идущим подряд черным символам, принадлежат одному 0-циклу. Поэтому каждый блок содержит не более n черных символов и не менее одного белого; последний блок содержит не менее двух белых символов.

Блок назовем *плохим*, если в нем содержится n черных и один белый символ. Заметим, что подряд может идти не более $n-2$ плохих блоков. Действительно, $n-1$ подряд расположенных плохих блоков образуют 1-цикл. Но тогда перестановки, соответствующие первому символу этого цикла и символу, непосредственно следующему за этим циклом, совпадают.

Итак, перед каждым хорошим блоком имеется не более $n-2$ плохих. Назовем *суперблоком* множество символов, состоящее из нескольких хороших блоков, перестановки которых принадлежат одному и тому же 0-циклу, и всех плохих блоков, непосредственно предшествующих им. Таким образом, все символы последовательности W, кроме красных, оказываются разбиты на суперблоки.

Пусть в некотором суперблоке имеется k хороших и l плохих блоков. Ввиду вышесказанного выполняется неравенство $l \leqslant k(n-2)$. Суперблок содержит $(l+1)n$ черных символов. Если $k = 1$, то в единственном хорошем блоке имеется n черных символов и, следовательно, не менее двух белых. Поэтому общее число белых символов не менее $l+2$. Если $k > 1$, то имеется не менее $l + k$ белых символов.

Пусть в суперблоке отношение числа белых символов к числу черных равно d/n. Оценим снизу величину d. Если $k = 1$, то

$$d \geqslant \frac{l+2}{l+1} = 1 + \frac{1}{l+1} \geqslant 1 + \frac{1}{n-1}.$$

Если $k \geqslant 2$, то

$$d \geqslant \frac{l+k}{l+1} = 1 + \frac{k-1}{l+1} \geqslant 1 + \frac{k-1}{1+k(n-2)} \geqslant 1 + \frac{1}{2n-3}.$$

Поскольку общее число черных символов равно $n!$, то общее число белых символов не менее

$$\frac{n!}{n}\Big(1 + \frac{1}{2n-3}\Big) = (n-1)! + \frac{n-1}{2n-3}(n-2)!$$

и длина последовательности не менее

$$n! + (n-1)! + \frac{n-1}{2n-3}(n-2)! + n - 3.$$

Теорема доказана.

3. Еще один метод построения той же последовательности

Второй метод построения универсальной последовательности использует метод построения последовательности слов из [3], где каждое очередное слово получено из предыдущего вставками блоков.

Для $n = 1, 2, \ldots$ определим последовательности X_n над алфавитом $\{1, \ldots, n\}$ следующим образом. Положим $X_1 = (1)$. Последовательность X_{n+1} строится из X_n вставками подслов по следующему правилу: для каждого подслова $x_i x_{i+1} \ldots x_{i+n-1}$ последовательности X_n, являющегося перестановкой символов $1, \ldots, n$, перед символом x_i вставляется слово $x_i x_{i+1} \ldots x_{i+n-1}(n+1)$. Например, $X_2 = (121)$, $X_3 = (123121321)$.

Теорема 4. *Последовательность X_n является универсальной для $P(n)$ и $|X_n| = n! + (n-1)! + \ldots + 2! + 1!$.*

Доказательство. Пусть $X_k = (x_1 \ldots x_p)$ и $X_{k+1} = (y_1 \ldots y_q)$. Последовательность X_{k+1} получается из X_k вставкой подслов длины $k+1$. Символы вставленных подслов назовем *новыми* в X_{k+1}, а символы из X_k назовем *старыми*.

Каждому старому символу y_j в X_{k+1} естественно соответствует некоторый символ x_i последовательности X_k. Докажем, что

$(*)$ *подслово $y_j \ldots y_{j+k-1}$ в X_{k+1} совпадает с подсловом $x_i \ldots x_{i+k-1}$ в X_k.*

Это очевидно, если не происходило вставок внутрь $x_i \dots x_{i+k-1}$. Если первая вставка произошла перед x_{i+l}, то по построению вставленное подслово начинается с $x_{i+l} \dots x_{i+k-1}$ и опять свойство $(*)$ доказано.

Докажем по индукции, что X_n имеет указанную длину и каждая перестановка из $P(n)$ встречается в X_n ровно один раз. При $n = 1, 2$ оба утверждения верны. Пусть они верны для X_k. Тогда последовательность X_{k+1} получается из X_k вставкой ровно $k!$ новых подслов длины $k + 1$, что доказывает первое утверждение. Далее, в каждом новом подслове символ $k + 1$ встречается один раз и подслова длины $2k + 1$ с центрами в этих символах имеют вид $w(k + 1)w$, где w пробегает по разу все перестановки множества $\{1, \dots, k\}$, — это следует из индукционного предположения и из доказанного выше свойства $(*)$. Итак, любое слово длины $k+1$, содержащее символ $k+1$, является перестановкой, все такие перестановки различны и среди них встречаются все перестановки из $P(k + 1)$. Теорема доказана.

Несложно показывается, что слова X_n, построенные выше, совпадают с универсальными словами $W_{n-2}(E)$, где E — тождественная перестановка.

В заключение отметим еще одно свойство построенных универсальных последовательностей. Из свойства $(*)$ следует, что те же самые последовательности X_n получаются, если использовать алгоритм вставки, зеркально симметричный приведенному: после каждой перестановки w длины k в X_k вставляется слово $(k + 1)w$. Отсюда следует, что слова X_n — палиндромы, т. е. одинаково читаются в обоих направлениях.

ЛИТЕРАТУРА

1. **Gallant J., Maier D., Storer J. A.** On finding minimal length superstrings // J. Comput. System Sci. 1980. V. 20, N 1. P. 50–58.
2. **Roth E.** Permutations arranged around a circle // Amer. Math. Monthly. 1971. V. 78, N 9. P. 990–992.
3. **Savage C.** A survey of combinatorial Gray codes // SIAM Rev. 1997. V. 39, N 4. P. 605–629.

Адрес авторов:

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

Статья поступила
20 января 2000 г.