УДК 51:621.396.801

АВТОМАТНЫЕ ПОЛЯ

Ю.Г.Косарев

I. Один из известных путей обогащения приемов программирования - реализация программными средствами конечных автоматов и их совокупностей. Такие программно-реализуемые (ПР-) автоматы обладают некоторыми свойствами, отличающими их от других классов конечных автоматов. Так, если у обычного автомата таблица переходов и таблица выходов задаются, как правило, в виде схемы, спрятанной в глубине "черного ящика", то у ПР-автомата эти таблицы, а также текущее состояние могут храниться в памяти ЭВМ в явном виде. Доступность этих таблиц позволяет легко вносить в них изменения,т.е. строить автоматы с переменной структурой.Отпадает также надобность для определения текущего состояния автомата или приведения его в заданное состояние решать специальные диагностическую или установочную задачи [I].

Как средство программирования ПР-автоматы могут при помощи одной несложной программы реализовать различные алгоритмы, соответствующие содержимому таблиц переходов и выходов. Выходам ПР-автомата могут быть поставлены в соответствие некоторые стандартные процедуры. Эти процедуры могут выполнять как внешние, так и внутренние функции, например, изменять таблицы переходов и выходов, состояние автомата.

Основные затраты памяти для осуществления ПР-автомата приходятся на таблицы переходов и выходов. Интересны случаи, когда эти таблицы одни и те же для больших совокупностей ПР-автоматов. Тогда для реализации каждого из них достаточно выделить в памяти участок для хранения номера его текущего состояния и организовать единую систему адресации, чтобы находить по имени автомата соответствующий ему участок памяти.

Условимся называть такие однородные совокупности ПР-автоматов **автоматными полями**, или **А-полями**.

А-поле можно характеризовать затратами памяти u в битах для записи текущего состояния каждого из автоматов А-поля и эффективностью α, которая характеризует качество решения задачи с помощью А-поля. (Это качество, естественно, конкретизируется в каждом случае.)

Обычно α зависит от u, поэтому имеет смысл говорить об удельной эффективности $\beta = \alpha / u$.

Покажем некоторые особенности применения А-полей на примере решения задачи нахождения одиночных элементов в больших последовательностях.

2. В соответствии с [2] решение данной задачи заключается в следующем. Имеется последовательность из N элементов, объем которой на много превосходит объем оперативной памяти (ОП). Подавляющая часть элементов входит в эту последовательность ровно один раз. Чтобы выявить повторяющиеся элементы, последовательными просмотрами исключают одиночные элементы. При каждом просмотре принимаются во внимание элементы, которым соответствует заданное значение ассоциативного кода z. Для этих элементов образуется другой ассоциативный код y, который принимается за номер автомата в А-поле.

Для различения элементов с одинаковыми кодами y образуется третий ассоциативный код x. (Все три кода x, y и z попарно независимы.) Требуется минимизировать число просмотров, что приводит к нахождению максимального значения коэффициента β, т.е. нахождению такого А-поля, для которого при каждом просмотре определяется наибольшее число одиночных элементов.

Более конкретно задача заключается в том, чтобы для заданного объема u, отведенного для каждого из автоматов А-поля, найти оптимальные таблицы переходов и выходов, а затем найти оптимальное значение u, при котором коэффициент β максимален.

3. Вероятностная модель. По условию задачи доля неодиночных элементов пренебрежимо мала. Распределение элементов по автоматам происходит с помощью ассоциативного кодирования. Это позволяет считать, что элементы равновероятно попадают в лю-

бой из автоматов. То есть данная задача совпадает с классической вероятностной задачей о распределении q дробинок по Q ящикам.

Согласно [3] математическое ожидание числа ящиков (автоматов) μ_r, в которые попало по r ($r = 0, 1, \ldots, q$) дробинок (элементов),

$$M\mu_r = C_q^r \frac{1}{Q^{r-1}} \left(1 - \frac{1}{Q}\right)^{q-r}. \qquad (1)$$

В нашей задаче q и Q велики, поэтому можно воспользоваться более простым выражением

$$M\mu_r = Q k^r e^{-k} / r!, \qquad (2)$$

справедливы при $q, Q \to \infty$ и при ограниченном $k = q/Q$. Нас интересует число выявленных одиночных элементов. Такими будут все элементы, которым соответствует $r = 1$, и часть элементов с $r = 2, 3, \ldots$, которые удается разделить с помощью ассоциативного кода x. Естественно, что разные автоматы могут обладать различными разделяющими способностями. Обозначим через p_r вероятность того, что все r наложившихся элементов разделятся. Тогда математическое ожидание числа одиночных элементов, выявленных при наложениях кратности r, можно представить в виде

$$h_r = r p_r M\mu_r. \qquad (3)$$

В данной задаче за величину коэффициента эффективности α можно взять среднее (по автоматам А-поля) число одиночных элементов, выявляемых за один просмотр:

$$\alpha = \frac{1}{Q} \sum_{r=1}^{q} h_r. \qquad (4)$$

Удельный коэффициент эффективности β, который характеризует число выявляемых за просмотр одиночных элементов, приходящихся на один бит памяти, занятой под А-поле, составит величину

$$\beta = \frac{\alpha}{u} = \frac{1}{Q} \sum_{r=1}^{q} h_r. \qquad (5)$$

Здесь $Qu = S$ – объем памяти под А-поле в битах. С учетом (2) и (3)

$$\beta = \frac{e^{-k}}{u} \sum_{r=1}^{q} p_r \frac{k^r}{(r-1)!}. \qquad (6)$$

Задача заключается в нахождении А-поля с наибольшим значением этого коэффициента.

Величина β зависит прежде всего от u. С ростом u растет возможное число состояний автомата $L = 2^u$, а следовательно, и возможность разделения групп элементов с большим значением r. (С увеличением L растет число членов в (6), для которых $\rho_r \neq 0$.) С изменением набора $\{\rho_r\}$ изменяется и оптимальное значение k, которое определяется из уравнения:

$$\sum_{r=1}^{q} \rho_r \frac{r}{r!} k^{r-1} (r-k) = 0, \qquad (7)$$

получающегося дифференцированием (6) по k.

4. Рассмотрим варианты А-полей для решения задачи выявления одиночных элементов. При этом будем предполагать, что для каждого варианта А-поля для размещения текущих состояний автомата выделено в ОП Q перенумерованных участков по u бит каждый и указана процедура, с помощью которой по номеру автомата определяется адрес его участка (ячейка и разряды в ячейке).

4.I. Можно видеть, что минимальное число состояний автомата для решений данной задачи равно трем: начальное E_0, конечное E_k при неуспехе и E_1, соответствующее поступлению на вход автомата одного кода.

В этом варианте $u = 2$, $\rho_1 = 1$, $\rho_i = 0$, $i > 1$. Оптимальные значения k и β согласно (6) и (7) равны:

$$k_{opt} = 1; \quad \beta_{opt} = \frac{1}{2e} = 0,184. \qquad (8)$$

4.2. Если использовать четвертое состояние автомата для частичного разделения пар (одной из четырех возможных) наложившихся элементов (рис.I), то $\rho_1 = 1$, $\rho_2 = \frac{1}{4}$, $\rho_i = 0$ $(i > 2)$ и соответственно:

$$k_{opt} = 1,24; \quad \beta = \frac{e^{-1,24}}{2} 1,24(1 + 0,31) = 0,236, \qquad (9)$$

что почти на 30% лучше (8).

4.3. В случае $u = 3$ (рис.2) удается разделить все неналожившиеся элементы ($\rho_1 = $ I); вс двойные, различающиеся

V_i	V_{i+1}	
	x=0	x=I
00	IO	II
OI	OI	OI
IO	OI	II
II	OI	OI

V	I(V)
00	0
OI	0
IO	I или 2
II	I

$K_{opt}=1,24$, $\beta=0,236$

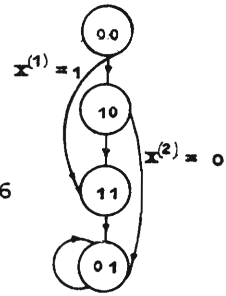
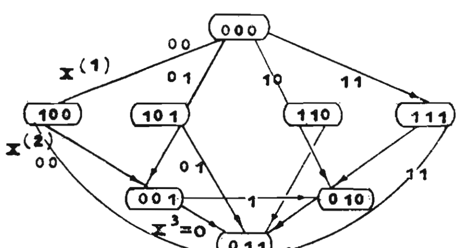


Рис.I. Таблица переходов $V_{i+1}(V_i\ x)$,таблица интерпретаций $I(V)$ и граф-схема переходов для случая $u=2$, $I(V)$ - число выявленных элементов, V - состояние автомата после просмотра всех элементов.

V_i	V_{i+1}			
	x=00	OI	IO	II
000	IOO	IOI	IIO	III
OOI	OII	OII	OIO	OIO
OIO	OII	OII	OII	OII
OII	OII	OII	OII	OII
IOO	OII	OOI	OIO	OIO
IOI	OOI	OII	OIO	OIO
IIO	OIO	OIO	OII	OIO
III	OIO	OIO	OIO	OII

V	I
000	0
OOI	2
OIO	2 или 3
OII	0
IOO	I
IOI	I
IIO	I
III	I



$K_{opt}=I,56$, $\beta=0,245$

Рис. 2. Таблицы $V_{i+1}(V_i,x), I(V)$ и граф-схема переходов для случая $u=3$.

двумя разрядами кода x, $(p_2 = \frac{3}{4})$ и часть тройных, соответствующих четырем комбинациям трех последовательных значений x : 00, 0I, I0; 00, 0I, II; 0I, 00, I0; 0I, 00, II $p_3 = \frac{1}{16}$.

В соответствии с (7)

$$k_{opt} = 1{,}56\,,$$
$$\beta_{opt} = \frac{e^{-1{,}56}}{3}\, 1{,}56(1 + 1{,}17 + 0{,}08) = 0{,}245. \quad (10)$$

При рассмотрении некоторых других вариантов организации автоматов с $u \geqslant 3$ значения β оказались меньше, чем в (I0).

4.4. Применение этого А-поля позволяет за один просмотр последовательности выявить $N_1 = 0{,}245\,S$ одиночных элементов. Для машины "Минск-32", если занять под А-поле половину ОП (то есть $S \approx 6 \cdot 10^5$ бит), то можно за один просмотр выявлять около $1{,}5 \cdot 10^5$ одиночных элементов из $2{,}3 \cdot 10^5$.

Какие именно элементы выявлены в действительности, становится известным после окончания просмотра, поэтому для выделения этих элементов нужен второй просмотр. Этот просмотр может быть совмещен с обработкой новой группы элементов [2]. Для этого каждый автомат нужно снабдить меткой, т.е. увеличить размер участка с 3 до 4 бит, что уменьшит N_1 до $1{,}1 \cdot 10^5$.

5. Можно видеть, что А-поля для решения данной задачи обладают следующими свойствами:

- все автоматы А-поля попарно независимы (т.е. выходы автоматов не соединены со входами других автоматов из А-поля);
- для интерпретации результатов каждому состоянию автомата, в котором он оказался после поступления на его вход анализируемой конечной последовательности кодов, с помощью таблицы интерпретаций ставится в соответствие значение некоторой функции;
- входной код либо не изменяет состояния автомата, либо переводит его в новое состояние, которое отличается от всех предыдущих состояний автомата;
- имеется начальное состояние E_0 автомата, которое он сохраняет до поступления на его вход хотя бы одного (любого) кода;
- имеются конечные состояния E_k автомата, которые сохраняются независимо от значения входного кода;

– имеются состояния E_i, пребывание в которых после завершения просмотра свидетельствуют об успехе.

В данном случае автоматы и работа с ними оказались довольно простыми. При необходимости для решения более сложных задач можно ввести особые состояния автомата E_c, в которых определенные входные коды могут изменять таблицу переходов или таблицу интерпретаций, число состояний и место размещения (и, естественно, адрес) автомата.

Литература

1. ГИЛЛ А. Введение в теорию конечных автоматов. Наука, М., 1966.

2. ГУСЕВ В.Д., КОСАРЕВ Ю.Г., ТИТКОВА Т.Н. Отыскание статистических закономерностей текстов методом ассоциативного кодирования. –Настоящий сборник, с. 72–89.

3. СЕВАСТЬЯНОВ Б.А., ЧИСТЯКОВ В.П. Асимптотическая нормальность в классической задаче о дробинках. –"Теория вероятностей и ее применение". 1964, т.IX, вып. 2, с. 223–237.

Поступила в ред.-изд.отд.
12 июня 1974 года