УДК 621.321.152

Ю. Н. АЛЕКСАНДРОВ, канд. техн. наук, *Э. Н. КОРОЛЬ*,
Ю. Ю. МИЛОНОВ, *А. В. ТОВАРНИЦКИЙ*, канд. техн. наук

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ОДНОГО МЕТОДА СЖАТИЯ МАССИВОВ ЦЕЛОЧИСЛЕННЫХ ДАННЫХ

В связи с использованием ЭВМ в процессе обработки информации широкое внедрение получают методы кодирования и декодирования таких совокупностей, как объективно-характеристические матрицы, различные монотонные последовательности, деструктизированные функции с ограниченным приращением, последовательности с ограниченной суммой компонент и т. д. [1].

В статье оценивается возможность использования метода полиадических чисел (МПЧ) в процессе сжатия информации, как одного из возможных методов решения подобного рода задач для сжатия массивов целочисленных данных.

Известно, что всякий целочисленный массив, представленный в виде матрицы $A = (a_{ij})$, $i = \overline{1, m}$; $j = \overline{1, n}$, с помощью преобразования $N_j = \sum_{i=1}^{m} a_{ij} P_i$ (1), где

$$P_i = \prod_{k=i+1}^{m} \lambda_k; \tag{2}$$

$$\lambda_i = 1 + \max_j (a_{ij}), \tag{3}$$

можно заменить двумя векторами $N = (N_j)$ и $\Lambda = (\lambda_i)$. При этом существует обратное преобразование:

$$A = (a_{ij}) = [N_j/P_i] - [N_j/\lambda_i P_i]\, \lambda_i, \tag{4}$$

позволяющее однозначно по компонентам векторов N, Λ восстановить любой элемент с погрешностью $\varepsilon = 0$ [2].

Оценку эффективности данного метода сжатия целочисленных массивов производим, используя соотношение [3], которое указывает, что объем двоичной памяти, занимаемый одним элементом при произвольном кодировании, $S_э = [\log_2 P]$ (5), где P — алфавит, применяемый для кодирования исходного информационного элемента. Объем памяти, необходимый для массива A размерности $m \times n$, равен

$$S_1 = mnS_э = mn\,[\log_2 P]. \tag{6}$$

При использовании же для представления информации на уровне элементов массивов обобщенного кода Бодо, построенного на основе метода упаковки по максимальному элементу, объем памяти, занимаемый одним информационным элементом, определяется соотношением [2]

$$S_{\text{ЭБ}} = [\log_2 \max_{i,j} a_{ij}]. \tag{7}$$

В этом случае объем памяти, требующийся для размещения массива размерности $m \times n$,

$$S_{\text{Б}} = mn\,[\log_2 \max_{i,j} (a_{ij})]. \tag{8}$$

С учетом соотношений (3), (5), (6) можно получить выражение для значения объема, занимаемого векторами Λ, N для первого варианта кодирования:

$$S_{\Lambda} = m\,[\log_2 (1 + \max_{j} (a_{ij}))]; \tag{9}$$

$$S_N = n\left[\log_2 \left(\sum_{i=1}^{m} a_{ij} \prod_{k=i+1}^{m} (1 + \max_{j} (a_{ij}))\right)\right] \tag{10}$$

и объем памяти, необходимый в этом случае для представления векторов N, Λ, которые характеризуют сжатый массив данных по методу полиадических чисел,

$$S^{1}_{\text{МПЧ}} = S_N + S_{\Lambda} = mS_{\text{Э}\Lambda} + nS_{\text{Э}N}. \tag{11}$$

С помощью соотношений (1), (2), (7), (8) можно определить объем памяти для представления векторов N и Λ в случае использования метода упаковки по максимальному элементу

$$S_{\Lambda} = [\log_2 \max_{i} (1 + \max_{j} (a_{ij}))]\, m; \tag{12}$$

$$S_N = nS_{\text{Э}N} = n\left[\log_2 \max_{j} \left(\sum_{i=1}^{m} a_{ij} \prod_{n=i+1}^{m} (1 + \max_{i} (a_{ij}))\right)\right], \tag{13}$$

а суммарный объем памяти, необходимый в таком случае для представления массива A, определяется как

$$\begin{aligned} S^{2}_{\text{МПЧ}} &= S_N + S_{\Lambda} = mS_{\text{Э}\Lambda} + nS_{\text{Э}N} = \\ &= n\left[\log_2 \max_{j} \left(\sum_{i=1}^{m} a_{ij} \prod_{k=i+1}^{m} (1 + \max_{j} (a_{ij})\right)\right] + \\ &\quad + m\,[\log_2 \max_{i} (1 + \max_{j} (a_{ij}))]. \end{aligned} \tag{14}$$

Оценку эффективности предложенного метода сжатия информации осуществим, используя традиционные для этого коэффициенты [4]:

$$\eta_i = S/S^{i}_{\text{МПЧ}}, \quad \beta_i = S_{\text{Б}}/S^{i}_{\text{МПЧ}} \quad (i = 1,\ 2),$$

которые с учетом выражений (5), (6), (8), (11) и (14) можно представить соотношениями

$$\eta_1 = \frac{mn\lceil \log_2 a_{ij}\rceil}{m\lceil \log_2 (1+\max\limits_{j}(a_{ij}))\rceil + n\lceil \log_2 \sum\limits_{i=1}^{m} a_{ij} \prod\limits_{k=i+1}^{m} (1+\max\limits_{j}(a_{ij}))\rceil}; \qquad (15)$$

$$\eta_2 = \frac{mn\lceil \log_2 a_{ij}\rceil}{n\lceil \log_2 (1+\max\limits_{j}(a_{ij}))\rceil + n\lceil \log_2 (\sum\limits_{i=1}^{m} a_{ij} \prod\limits_{k=i+1}^{m} (1+\max\limits_{j}(a_{ij}))\rceil}; \qquad (16)$$

$$\beta_1 = \frac{mn\lceil \log_2 \max\limits_{ij}(a_{ij})\rceil}{m\lceil \log_2 (1+\max\limits_{j}(a_{ij})\rceil + n\lceil \log_2 (\sum\limits_{i=1}^{m} a_{ij} \prod\limits_{k=i+1}^{m} (1+\max\limits_{j}(a_{ij}))\rceil}; \qquad (17)$$

$$\beta_2 = \frac{mn\lceil \log_2 \max\limits_{ij}(a_{ij})\rceil}{m\lceil \log_2 \max\limits_{i}(1+\max\limits_{j}(a_{ij}))\rceil + n\lceil \log_2 \max\limits_{j}(\sum\limits_{i=1}^{m} a_{ij} \prod\limits_{k=i+1}^{m} (1+\max\limits_{i}(a_{ij})))\rceil}; \qquad (18)$$

где η_1, η_2 — коэффициенты, характеризующие выигрыш в количестве разрядов представления, получаемый за счёт сжатия информации (МПЧ), которая в исходном виде была представлена равномерным кодом, по отношению к количеству разрядов записи сжатых массивов соответственно равномерным кодом и кодом Бодо; β_1, β_2 — коэффициенты, характеризующие выигрыш в количестве разрядов представления, получаемый в результате применения при сжатии информации МПЧ, которая в исходном виде была представлена кодом Бодо, по отношению к количеству разрядов записи сжатых массивов равномерным кодом и кодом Бодо.

Как показывает анализ соотношений (15) — (18), для конкретных целочисленных массивов $A = (a_{ij})$ размерности $m \times n$ ($m \geqslant 5$; $n \geqslant 5$) выигрыш от использования МПЧ при сжатии информации составляет величину более 1,7 раза, что говорит о целесообразности технической разработки данного метода сжатия информации. С этих позиций более предпочтительным, но менее конструктивным является представление сжатых массивов с помощью кода Бодо. Предметом дальнейших исследований для увеличения коэффициентов η_i, β_i следует считать определение кратности применения МПЧ к множеству векторов Λ и N, на основе которых можно сформировать очередной (новый) массив.

Список литературы: 1. *Кожурин Ф. Д., Ярмош Н. А.* Структурная обработка больших информационных массивов. Минск, 1973. 256 с. 2. *Амелькин В. А.* О нумерических функциях для кодирования числовых данных // Применение математических методов и ЭВМ при поиске полезных ископаемых. Новосибирск, 1986. С. 89—93. 3. *Акушский И. Я., Заблоцкий Б. Н.* О комбинаторном подходе к идее сжатия информации // Цифровая вычислительная техника и программирование. М., 1971. Вып. 6. С. 5—17. 4. *Стахов А. П.* Введение в алгоритмическую теорию измерений. М., 1977. 173 с.

Поступила в редколлегию 26.08.88