

В. Лужецький, В. Каплун (м. Вінниця)
МЕТОДИ УЩІЛЬНЕННЯ ЧИСЛОВИХ
ПОСЛІДОВНОСТЕЙ, ЩО БАЗУЮТЬСЯ
НА ВІДХИЛЕННІ ВІД КОНСТАНТ

Відомі методи ущільнення базуються або на статистичних характеристиках даних, що підлягають ущільненню, або використовують словниковий принцип. Але, оскільки сучасні мікропроцесори найефективніше здійснюють арифметичні операції над числами, то останнім часом відбуваються пошуки підходів щодо ущільнення даних, які базуються на їх представленні як цілих чисел. У доповіді розглядаються методи ущільнення даних, основою яких є обчислення відхилень від елементів, вибраних за певними правилами і таких, що не враховують статистичних характеристик. Одним з правил, що використовується в запропонованих методах, є введення констант.

Нехай початкова послідовність n -розрядних чисел Q складається з K елементів q_i : $Q = \{q_1, q_2, \dots, q_k\}$. У цій послідовності мінімальне значення чисел q_{\min} , максимальне — q_{\max} ; діапазон значень $D = q_{\max} - q_{\min}$.

Коли використовується одна константа C , то вона розташовується таким чином, щоб максимальне значення відхилень чисел від константи були однаковими для чисел, що менше за константу, і чисел, що більші за константу. Отже,

$$C - q_{\min} = q_{\max} - C, \text{ або } 2C = q_{\max} - q_{\min}, \text{ звідки } C = D/2.$$

В результаті обчислення відхилень отримуємо послідовність:

$$Q_{\text{ущ.}} = \{d_1; d_2; \dots; d_k\}, \text{ де } d_i = q_i - C, i = \overline{1 \div K}.$$

В ущільненій послідовності $Q_{\text{ущ.}}$ кожний елемент вхідної послідовності Q представляється значенням його відхилення від вибраної константи. Серед отриманих відхилень будуть як додатні, так і від'ємні числа. Тому, якщо запам'ятовувати знак z_i в окремому розряді і відкидати нульові старші розряди для додатних чисел і одиничні — для від'ємних, запам'ятовуючи їх кількість v_i , то залишаються числа d_i^* , які матимуть різну розрядність. Тоді виникає задача їхнього відокремлення для можливості подальшого однозначного відновлення. Структура ущільненої інформації буде такою:

d_1^*	d_2^*	...	d_K^*	v_1	v_2	...	v_K	z_1	z_2	...	z_K
---------	---------	-----	---------	-------	-------	-----	-------	-------	-------	-----	-------

При даному підході ущільнення досягається за рахунок відкидання розрядів, оскільки замість l відкинутих розрядів ($l = 1 \div n - 1$) записується код v_i розрядності $\lceil \log_2(n-1) \rceil$.

При використанні двох констант C_1 і C_2 вони розташовуються таким чином, щоб відхилення чисел послідовності від них було найменшим. При цьому можливі декілька випадків розташування констант, але при будь-якому варіанті їх розташування спостерігаються певні тенденції.

1. Якщо константи і правило отримання відхилень вибирати так, щоб відхилення були одного знаку, то розрядність відхилень буде меншою, але треба мати один розряд на кожне відхилення для уточнення його розташування відносно константи (ознака константи).

2. Якщо константи і правило отримання відхилень вибирати, не зважаючи на знак відхилення, то можна зекономити на кількості відкинутих розрядів, але треба зберігати дві ознаки (одна – для знаку різниці, друга – ознака константи). Отже, немає необхідності штучно вибирати розташування констант – головне, щоб вони були рівномірно розкидані по діапазону чисел вхідної послідовності.

У загальному випадку при виборі N констант за правилом:

$$C_i = \frac{(2i-1) \cdot D}{2N}, i = 1, 2, \dots, N,$$

кожному елементу з вхідної послідовності Q в ущільненій послідовності $Q_{\text{ущ}}$ відповідатиме набір із значення відхилення і ознаки константи c_i (для $N > 1$), за допомогою якої воно обчислювалось. При цьому кожне відхилення буде представлятися сукупністю таких елементів: знака відхилення z_i ; кількості відкинутих розрядів v_i у значенні різниці і числа d_i^* , яке залишилося.

Таким чином, структура ущільненої інформації буде такою:

d_1^*	...	d_K^*	v_1	...	v_K	z_1	...	z_K	c_1	...	c_K
---------	-----	---------	-------	-----	-------	-------	-----	-------	-------	-----	-------

Найбільший коефіцієнт ущільнення для запропонованих методів досягається, коли значення чисел вхідної послідовності Q розподілені за нормальним законом із центрами, значення яких дорівнюють значенням введених констант C_i . Оскільки реальні послідовності Q можуть відповідати іншим законам розподілу значень, то для досягнення найбільшого коефіцієнта ущільнення потрібно здійснювати перетворення послідовності Q .