УДК 519.8

Я.А. Кардаш, Г.Г. Цегелик

ДО ПОБУДОВИ ОПТИМАЛЬНИХ ІМОВІРНІСНИХ ДВІЙКОВИХ ДЕРЕВ ПОШУКУ

Постановка задачі. Дано n значень ключа $K_1, K_2, ..., K_n$ ($K_1 < K_2 < ... < K_n$) і $2n+1$ чисел $p_1, p_2, \dots p_n$, $q_0, q_1, \dots q_n$, які задовольняють умову $p_1 + p_2 + \dots + p_n + q_0 + q_1 + \dots + q_n = 1$, де p_i – імовірність того, що аргументом пошуку є K_i; q_j – імовірність того, що аргумент пошуку лежить між K_j і K_{j+1} (q_0 – імовірність того, що аргумент пошуку менший K_1, а q_n – імовірність того, що аргумент пошуку більший K_n). Потрібно побудувати двійкове дерево із значеннями вершин $K_1, K_2, ..., K_n$, для якого досягається мінімум математичного сподівання числа порівнянь, необхідних для пошуку вершини,

$$E = \sum_{i=1}^{n} l_i p_i + \sum_{j=0}^{n} s_j q_j ,$$

де l_i – номер рівня, на якому розташована вершина із значенням K_i, s_j – більший з номерів рівня, на якому міститься вершини з номерами K_j та K_{j+1}.

Математичне сподівання числа порівнянь, необхідних для пошуку, прийнято називати ціною дерева, а дерево з мінімальною ціною – оптимальним деревом. При такому визначенні не обов'язково вимагати, щоб $p_1 + p_2 + \dots + p_n + q_0 + q_1 + \dots + q_n = 1$, можна шукати дерево з мінімальною ціною для заданої послідовності „ваг" ($p_1, p_2, \dots p_n$, $q_0, q_1, \dots q_n$).

В [1] наведений алгоритм побудови оптимального дерева, трудомісткість якого $O(n^3)$. Така трудомісткість зумовлена тим, що в основі цього алгоритму лежить фактично перебір варіантів. В даній роботі для розв'язання задачі пропонується використати співвідношення оптимізації як це було зроблено в [2, 3] у випадку $q_0 = q_1 = \dots = q_n = 0$.

Позначимо через $S_r(i,j)$ ціну піддерева з вагами $(p_{i+1}, p_{i+2}, \dots, p_j, q_i, q_{i+1}, \dots, q_j)$ і коренем K_r, де $0 \le i \le r \le j \le n$, а через $\omega(i,j)$ – суму цих ваг. Введемо величини:

$$S_r^+(i,j) = \sum_{k=i+1}^{j} (l_k + 1) p_k + \sum_{k=i}^{j} (s_k + 1)\, q_k ,$$

$$S_r^-(i,j) = \sum_{k=i+1}^{j} (l_k - 1) p_k + \sum_{k=i}^{j} (s_k - 1)\, q_k .$$

Тоді

$$S_r^+(i,j) = S_r(i,j) + \omega(i,j),$$

$$S_r^-(i,j) = S_r(i,j) - \omega(i,j).$$

Звідси

$$\omega(i,j) = S_r^+(i,j) - S_r(i,j) = S_r(i,j) - S_r^-(i,j).$$

Одержана формула використовується для виведення так званих співвідношень оптимізації, тобто умов, накладених на ваги, при яких перехід від піддерева з коренем A (рис. 1) до піддерев з коренями B, C, E або F не веде до зменшення ціни піддерева.

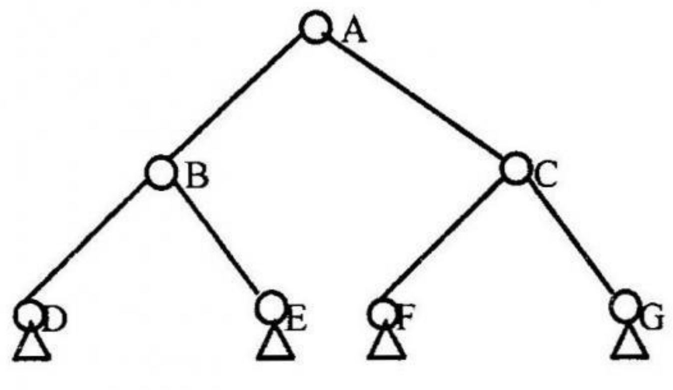

Рис. 1

Маючи ці умови, будується алгоритм оптимізації дерева.

1. Кнут Д. Искусство программирования для ЭВМ: В 3 т. Т.3: Сортировка и поиск - М.: Мир, 1978. - 844с. 2. Цегелик Г.Г. Методы автоматической обработки информации. - Львов: Вища школа, 1981. - 132с. 3. Цегелик Г.Г. Алгоритм оптимизации двоичных вероятносных поисковых деревьев. - В сб.: Исследование операций и АСУ. 1980. № 16. С.78-82.