談 話 室

2分表引きの探索回数について*

大 駒 誠 一**

計算機のプログラムで順序のそろった大きな表を引くのに，2分表引き法 (binary search) を使うのが常套手段であるが，その効率の計算に関しては本誌の記事も含めてかなり誤解があるように思われる．

たとえば，表の要素の数を n 個として，その中の特定の要素を2分表引きで探し出すまでに比較をする回数は，本によって次のようにまちまちで，いずれも正しくない．

a）たかだか　$\log_2 n$ 回

d）平　　均　$\log_2 n$ 回

c）最　　大　$[\log_2 n]$ 回

b）だいたい　$\log_2 n$ 回†

昨年行なわれた情報処理技術者認定試験（問題1の前の問 32）に2分表引きの最大比較回数を求める問題があったが，その後たくさん出た解答例のうち，約半数は答が間違っていた．

要素番号	1	2	3	4	5	6	7	8
比較回数	3	2	3	1	3	2	3	4

第1図　要素番号と比較回数（$n=8$ の場合）

比較回数をくわしく計算してみよう．簡単のために n がちょうど 2^k（k は自然数）の場合を考える．第1図を見てもわかるとおり，どちらかの一番はしの要素を検出するためには $k+1$ 回比較しなければならない．平均比較回数を求めるために，1，2，3……回の比較でみつかる要素の数がそれぞれ何個あるか調べてみると

1回でみつかる要素	1個
2回でみつかる要素	2個
3回でみつかる要素	4個
⋮	
k 回でみつかる要素	2^{k-1} 個
$k+1$ 回でみつかる要素	1個

したがって平均比較回数 S は

$$S=\{1\times1+2\times2+3\times4+\cdots\cdots+k\times2^{k-1}+(k+1)\times1\}/n$$
$$=\{2^k(k-1)+(k+2)\}/n \qquad (1)$$

ここで，$n=2^k$，$k=\log_2 n$ だから

$$S=\log_2 n-1+\frac{\log_2 n+2}{n} \qquad (2)$$

となる．n がある程度大きければ最後の項は無視できるので，平均比較回数は約 $\log_2 n-1$ 回ということができる．

(2)式は n がちょうど 2^k でないときにもだいたいあてはまる．したがって，2分表引きの比較回数は，一般に

最大　$[\log_2 n]+1$ 回

平均　約 $\log_2 n-1$ 回

である．

これをみてもわかるように，最大と平均との差があまりないために，誤解を生じているのかもしれない．

以上の話は，キーの分布が一様分布であることを仮定しているのはいうまでもない．

（昭和 45 年 3 月 3 日受付）

* On Efficieney of Binary Search, by Seiichi Ohkoma (The Faculty of Engineering, Keio University)

** 慶応塾義大学工学部

† 「だいたい」なのでこれは間違いとはいえない