Selection in worst-case linear time

(OR)

Median-finding algorithm

(OR)

Median-of-medians algorithm

⇒ The algorithm "DETERMINISTIC-SELECT" has the running time $O(n)$ in the worst case. Like RANDOMIZED-SELECT, this algorithm finds the desired element by recursively partitioning the input array.

⇒ Here, we guarantee a good split upon partitioning the array.

⇒ The algorithm uses the PARTITION algorithm of QUICKSORT (but modified to take the element to partition around as an input parameter).

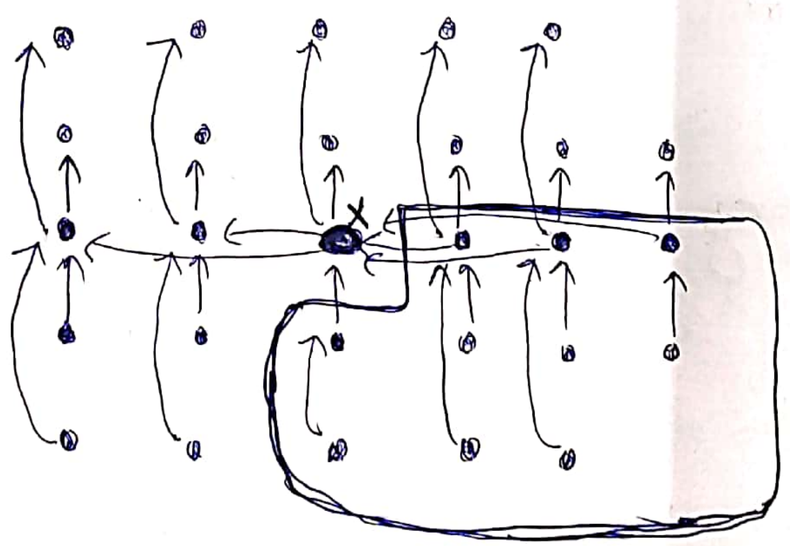
DETERMINISTIC-SELECT (A, p, r, i)

1) Divide the n elements of the input array into $\lfloor \frac{n}{5} \rfloor$ groups of 5 elements each and at most one group made up of the remaining n mod 5 elements.
2) Find the median of each of the $\lceil \frac{n}{5} \rceil$ groups by first insertion-sorting the elements of each group and then picking the median from the sorted list of group elements.
3) Apply DETERMINISTIC-SELECT recursively to find the median (x) of $\lceil \frac{n}{5} \rceil$ medians found in step (2).

4) Partition the input array around the median-of medians "x" using the modified version of PARTITION. So, x is the k^{th} smallest element (where, k is one more than the number of elements on the low side (left side) of the partition, and there are $(n-k)$ elements on the right side of the partition.

5) If $(i==k)$, then return x. Otherwise, use DETERMINISTIC-SELECT recursively to find the i^{th} smallest element on the left side (if $i<k$); or find the $(i-k)^{th}$ smallest element on the right side (if $i>k$).

Analysis of Running Time :-



(Arrows go from larger elements to smaller elements)

⇒ Determine a lower bound on the number of elements that are greater than X i.e. find atleast how many elements are greater than x.

Scanned with CamScanner

⇒ From the figure, we can visualize that At least half of the medians (found in step (2)) are greater than or equal to the medians-of-medians "x".

⇒ Thus, at least half of the $\lceil \frac{n}{5} \rceil$ groups contribute at least 3 elements that are greater than x; (except the group that has fewer than 5 elements and the group which contains x itself).

⇒ Number of elements greater than x is at least $\Big\}$ $3\left(\left\lceil \frac{1}{2}\left\lceil \frac{n}{5} \right\rceil \right\rceil - 2\right)$

(Discounting incomplete group and the group that contains x)

$\geq \frac{3n}{10} - 6$.

So, the number of elements greater than x $\Big\}$ is at least $\frac{3n}{10} - 6$.

⇒ So, in the worst-case step(5) calls DETERMINISTIC-SELECT recursively on at most $\frac{7n}{10} + 6$ elements.

$$T(n) \leq T\left(\left\lceil \frac{n}{5} \right\rceil\right) + T\left(\frac{7n}{10} + 6\right) + O(n)$$

(step (3)) (step (5)) (O(n) calls of INSERTION SORT on lists of size O(1))

$$\leq c\left\lceil \frac{n}{5} \right\rceil + c\left(\frac{7n}{10} + 6\right) + an$$

$$\leq \frac{cn}{5} + c + \frac{7cn}{10} + 6c + an$$

$$= \frac{9cn}{10} + 7c + an$$

$$= cn - \frac{cn}{10} + 7c + an$$

$$= cn + \left(-\frac{cn}{10} + 7c + an\right)$$

$$\Rightarrow T(n) \leq cn + \left(-\frac{cn}{10} + 7c + an\right)$$

which is at most cn if;

$$-\frac{cn}{10} + 7c + an \leq 0$$

On solving we get;

$$c \geq 10a\left(\frac{n}{n-70}\right)$$

So, n must be strictly greater than 70.

We take $n = 140$, and get $c \geq 20a$.

So, we can conclude that worst-case running time of DETERMINISTIC-SELECT is $O(n)$; and the recurrence relation is of the form :-

$$T(n) \leq \begin{cases} O(1) & \text{if } n < 140 \\ T(\lceil \frac{n}{5} \rceil) + T(\frac{7n}{10} + 6) + O(n) & \text{if } n \geq 140 \end{cases}$$

— × — × — × —

"Forming groups of 7"

⇒ Number of elements greater than x is at least $\left.\right\}$ $4\left(\left\lceil \frac{1}{2}\left\lceil \frac{n}{7} \right\rceil \right\rceil - 2\right)$

$$\geq \frac{4n}{14} - 8$$

$$= \frac{2n}{7} - 8.$$

So, the number of elements greater than x is at least $\frac{2n}{7} - 8$

→ So, in the worst-case step(5) calls the algorithm on at most $\frac{5n}{7} + 8$ elements.

$$T(n) \le T\left(\left\lceil \frac{n}{7} \right\rceil\right) + T\left(\frac{5n}{7} + 8\right) + O(n)$$

$$\le c\left\lceil \frac{n}{7} \right\rceil + c\left(\frac{5n}{7} + 8\right) + an$$

$$\le \frac{cn}{7} + c + \frac{5cn}{7} + 8c + an$$

$$= \frac{6cn}{7} + 9c + an$$

$$= cn - \frac{cn}{7} + 9c + an$$

$$T(n) \le cn + \left(-\frac{cn}{7} + 9c + an\right).$$

which is at most cn if ;

$$-\frac{cn}{7} + 9c + an \le 0.$$

On solving we get ;

$$c \ge \frac{7an}{n-56}$$

So, n must be strictly greater than 56.

We take $n = 112$; and get $c \ge 14a$

$$T(n) \le \begin{cases} O(1) & \text{if } n < 112 \\ T\left(\left\lceil \frac{n}{7} \right\rceil\right) + T\left(\frac{5n}{7} + 8\right) + O(n) & \text{if } n \ge 112 \end{cases}$$