

2-2 Treasure Hunting

(Monday, April 2, 2018 ~ Saturday, April 7, 2018)

Hengfeng Wei

hfwei@nju.edu.cn

April 2, 2018





Problem of the Week (Monday, April 2, 2018 ~ Saturday, April 7, 2018)

- (a) Given an array $A[0 \cdots n - 1]$, to determine whether there is a value that *occurs more than $\lfloor n/k \rfloor$ times* in $\Theta(n \lg k)$ time and $\Theta(k)$ extra space.
- (b) Prove that the *lower bound* of this problem is $\Theta(n \lg k)$.

TIP#1

(Monday, April 2, 2018)

Take $k = 2$.

$\Theta(n)$ time & $\Theta(1)$ space

TIP#2

(Tuesday, April 3, 2018)

Definition (k -simplified Multiset)

Consider a multiset \mathcal{M} . A *k -simplified multiset* for \mathcal{M} is a multiset derived from \mathcal{M} by repeating *deleting k distinct elements* from it until no longer possible.

Theorem

If a value occurs more than $\lfloor n/k \rfloor$ times in \mathcal{M} of n elements, then it is in a k -simplified multiset for \mathcal{M} .

Prove this theorem. Take $k = 2$ again. Design an $\Theta(n)$ algorithm for $k = 2$. Generalize it to an algorithm for general k (ignoring $\Theta(n \lg k)$ for now).

TIP#3

(Wednesday, April 4, 2018)

Today, you have an efficient data structure T for a multiset:

We denote a multiset by $\mathcal{M} = \{(v_i, c_i)\}$, where c_i is the number of times v_i occurs in \mathcal{M} . The number of distinct values in \mathcal{M} is denoted by $d = |\{v_i\}|$.

The data structure T for \mathcal{M} contains d nodes, each being a pair (v_i, c_i) . It supports INSERT, DELETE, and SEARCH in $\Theta(\lg d)$.

Use this data structure (you are not required to implement it) in your algorithm to achieve the time complexity of $\Theta(n \lg k)$.

TIP#4

(Thursday, April 5, 2018)

Definition (r -list)

Let $r = \lfloor n/k \rfloor$. An r -list is a list of n values such that each of the values $0, 1, \dots, \lfloor n/r \rfloor - 1$ occurs r times and the value $\lfloor n/r \rfloor$ occurs $(n \bmod r)$ times.

Lemma

There are at least $(k/e)^n$ different r -lists.

Theorem

Executing these r -lists on a decision tree will follow different paths.

TIP#5

(Friday, April 6, 2018)

Definition ($L = L_1 * L_2$)

Given two lists L_1 and L_2 , a new list $L = L_1 * L_2$ is defined as follows:

$$L[i] = \min \{L_1[i], L_2[i]\}, 0 \leq i < n$$

Theorem

*If r -lists L_1 and L_2 are different, then $L = L_1 * L_2$ **contains** a value which occurs more than r times.*

Prove the theorem on Thursday **by contradiction**,
with the help of the definition and theorem above.

TIP#6

(Saturday, April 7, 2018)

Paper ([Clickable](#)):

Finding Repeated Elements (J. Misra, David Gries, 1982)



Thank
You!



Office 302

Mailbox: H016

hfwei@nju.edu.cn