Middle 解题报告

杭州外国语学校　陈立杰

Contents

1 题目大意

一个长度为n的序列a,它的中位数$middle(a)$,定义为序列a,从小到大排序排序之后的序列b的第$\lfloor \frac{n}{2} \rfloor$位,序列的下标从0开始.

举例的话序列$middle\{2,1\}$就是排序后$\{1,2\}$的第1位即2.

现在给长度为n的序列s,定义$s[l,r]$为s的从第l位到第r位组成的连续子序列. 现在有q个询问,每次询问$l \in [a,b], r \in [c,d]$的$middle(s[l,r])$的最大值. 对所有询问,有$a < b < c < d$. 必须在线回答所有问题.

2 数据范围

- 5% $n, q \leq 100$
- 30% $n, q \leq 2000$
- 100% $n \leq 20000, q \leq 25000$

3 本文中需要的一些基础知识

- 排序 我相信参加NOI的选手应该都会使用一种$O(n \log n)$复杂度的排序,故不作介绍.
- 线段树 我相信参加NOI的选手应该都会使用线段树求最大子段和,最大前缀和,故不作介绍.
- 可持久化数据结构 这方面的内容比较新颖,我会在文中加以说明.

4 5%的做法

对于$n, q \leq 100$的数据,我们对于每个子区间$[l,r]$,通过排序求出它的中位数,询问也挨个询问所有的$l \in [a,b], r \in [c,d]$的区间即可. 复杂度是$O(n^3 \log n + qn^2)$.

5 进一步的分析

显然通过暴力算法是很难解决这个问题的,我们得分析一下中位数的性质.

我们不妨考虑二分答案,容易发现,一个长度为n的序列s,它的中位数$\geq x$的条件,就是$\geq x$的数的数量,大于等于$< x$的数的数量.

那么我们不妨将$\geq x$的数看成+1,$< x$的数看成-1,那么序列s的中位数$\geq x$的条件,就是转换之后该序列的和≥ 0.

进一步的,要判断$l \in [a,b], r \in [c,d]$的区间中,有没有中位数$\geq x$的,只需要判断转换之后,这些区间中有没有和≥ 0的.

这个只要求出$l \in [a,b], r \in [c,d]$的区间中和最大的,就可以判断了.

由于$a < b < c < d$,那么实际上这个区间可以看成$[l, b] + [b + 1, c - 1] + [c, r]$,也就是$[a, b]$的最大后缀和,$[b + 1, c - 1]$的和,$[c, d]$的最大前缀和.

6 20%的做法

根据前面的分析,我们在判断答案是否可能$\geq x$的时候,只需要将所有数重新标号成$+1, -1$,然后求几个区间的和或最大前缀或最大后缀和即可. 由于可能的答案只有$O(n)$个,我们对数列中的每个值,都重新标号之后使用线段树来维护.就能在$O(\log n)$的时间内判断答案是否可能$\geq x$,那么因为同时需要二分$O(\log n)$次,故每次询问复杂度为$O(\log^2 n)$.

总时间复杂度为$O(n^2 + q \log^2 n)$.

7 可持久化线段树

什么是可持久化数据结构呢,简单的说就是一种能够维护历史版本的数据结构.

让我们来考虑如何实现可持久化线段树.

首先我们的目标是,修改一个位置上的值,同时不影响所有目前保存的历史版本的运作,得到一个新的版本的线段树.

由于线段树的实现是递归的,那么我们也来递归的看待这个问题.

首先考虑一个叶子节点,那么我们只需要新建一个新的叶子节点,它的值是修改过后的值,那么就能得到一个当前版本.

再考虑一个非叶子节点,由于它的两个孩子中最多只会有一个被修改,不妨看成左孩子,那么我们对左孩子递归调用函数,得到左孩子的修改后版本,然后将当前节点拷贝一份,右孩子不变,左孩子为修改后的版本,就能得到当前节点修改后的版本.

注意到会被修改的节点,只能是被修改的叶子节点的祖先,故每次只需要新建$O(\log n)$个结点.

同时我们只是在新建节点,没有对任何节点的信息做修改,故历史版本也没有收到影响.

8 100%的算法

不妨将s中的数进行排序,那么关于s_i的线段树,和关于s_{i+1}的线段树,只有一个节点被修改了.

那么套用我们之前提到的可持久化线段树,只需要$O(n \log n)$的时间和空间,就能得到关于所有s_i的线段树.

那么总时间复杂度就是$O(n \log n + q \log^2 n)$.

可以圆满的解决问题.

9 一些题外话

可持久化线段树的功能还不仅于此,他本身就有离线变在线的特点,同时还有很多意想不

到的用处.

- $O(n\log n)+O(\log n)$的区间找第k小的数.
- $O(n\log n)+O(\log n)$的找树上一条路径上第k小的数.
- 上两个带修改的话可以得到$O(n\log^2 n)+O(\log^2 n)$的复杂度.

p 这个题目还有一种$n^{1.5}+q\sqrt{n}\log n$的算法,遗憾的是不能通过所有数据,只能得30分.

还有一种使用划分树的,复杂度也是$n\log n+q\log^2 n$的算法,不过常数和编写复杂度都较大,涉及的知识也更多,就略去不谈.

10 出题灵感

在刚刚过去的WC 2012上,fhq神犇讲了可持久化数据结构,我就想出一道使用这个的题目,经过一些酝酿,这题就应运而生了.