数据结构常见模型及应用

2024 年 7 月

feecle8146

回顾

昨天我们学习了树状数组、倍增、dfs 序三个主要技巧。试着回忆一下：

1. 树状数组的循环不变式是什么？
2. 树状数组如何维护区间加区间求和？如何拓展到二维？
3. 如何设计倍增信息？
4. dfs 序的作用是什么？如何使用 dfs 序维护到根加子树求和？

今天，我们将学习几个常见模型，在模型中运用之前学习的数据结构。

单调栈

单调栈的结构

单调栈回答了如下的问题：给定一个序列 a。对于每个 $1 \le i \le n$，维护出 $a[1 \dots i]$ 这个前缀的所有后缀最大值的信息。

思考：$a[1 \dots i]$ 的后缀最大值相比于 $a[1 \dots i-1]$ 的后缀最大值怎么改变？

只有 $\le a_i$（被 a_i 比下去的）会改变。

据此可以写出下面的代码：

```
int st[N + 5], top;

st[0] = top = 0; // 特别是多次单调栈不能忘！
for (int i = 1; i <= n; i++) {
    while (top && a[i] >= a[st[top]]) top--;
    st[++top] = i;
    // 此时 st 就存储了 [1..i] 的后缀最大值下标
}
```

单调栈

单调栈的结构

当然，对于前缀/后缀最小值，方法是一样的。如果元素有相同的，需要想清楚取不取等。

练习：

1. 证明前述代码时间复杂度为 $O(n)$。
2. 右侧代码的时间复杂度？

```
st[0] = top = 0;
for (int i = 1; i <= n; i++) {
    while (top && a[i] >= a[st[top]]) top--;
    st[++top] = i;
    for (int j = 1; j <= top; j++) {
        printf("%d\n", st[j]);
    }
}
```

由于所有前缀的所有后缀就是所有区间，所以我们已经获得了所有区间的最值信息，可惜不是显式的（或者说可枚举的）。

感兴趣的同学可以课后自学“笛卡尔树”。

单调栈

区间最大值 无来源

给你一个数组，多次询问区间最大值。

要求用单调栈解决。

单调栈在扫到 i 的时候，我们就获知了 $[1 \ldots i]$ 的所有后缀的最大值信息，也就是右端点为 i 的所有区间的答案。

然而，扫到 $i+1$ 的时候，单调栈就改变了。所以，只有扫到 i 这个时刻，才能回答 $[\ldots, i]$ 的询问。由此，你能想到解决办法吗？

离线询问，将询问 $[l, r]$ 放进 r 的 vector 里。对询问 $[l, r]$，在单调栈上二分出最小的 $\geq l$ 的后缀最大值。

单调栈

区间最大值之和 无来源

给你一个数组，求所有区间 $[l,r]$ 的最大值之和。

要求线性。

单调栈在扫到 i 的时候，我们就获知了 $[1\ldots i]$ 的所有后缀的最大值信息。

$[1\ldots i]$ 所有后缀的最大值之和是多少？

$$\sum_i a_{st_i} \times (st_i - st_{i-1})$$

可以在单调栈上维护前缀和。本做法其实对所有前缀 $a[1\ldots i]$ 都求了答案。

单调栈

区间后缀最大值和 无来源

给你一个数组，多次询问区间的所有后缀的最大值之和。

$$n, q \le 300000$$

单调栈在扫到 i 的时候，我们就获知了 $[1 \dots i]$ 的所有后缀的最大值信息，也就是右端点为 i 的所有区间的答案。

结合预处理的 $\sum_i a_{st_i} \times (st_i - st_{i-1})$，只需找出 l 在 st 上的分界点（$st_j < l \le st_{j+1}$），特殊处理一下 j 这一段，其它的就是前缀和相减。

单调栈

prev 无来源

给你一个数组，对于每个 i，求最大的 $j < i$ 使得 $a_j > a_i$。

要求线性。

就是 i 处单调栈（还没把 i 加进去时）的栈顶。

据此，能否用另一种方法解决前一题？

还可以类比：实数区间覆盖，只能用半开半闭区间才不重不漏。

记本题答案为 l_i，从右往左的答案为 r_i。i 作为区间最小值，几乎等价于 $l \in (l_i, i], r \in [i, r_i)$，但 l_i, r_i 需要一个取等一个不取等！

为了理解红色部分，可以考虑 $a = [1,1,1]$ 的情况。

单调栈

带删数的区间最大值之和 CF1988E

定义数组 a 所有区间的最大值之和为 $f(a)$。

设 a 删掉第 i 个元素得到的数组为 b_i，求所有 $f(b_i)$。

如何将上述问题的做法泛化到本题？

提示：枚举最小值位置 j，考虑 j 对哪些 b_i 有贡献。

分 i 所在区间（与 l_j,r_j 位置关系）讨论，用差分计算答案。

单调栈

体育馆问题

CF1601E

有一个体育馆，第 $i\,(1 \le i \le n)$ 天，票价为 a_i 元。每张你手上的票可以管任意连续的 k（定值）天，也就是说，如果你在第 i 天买了这张票，你可以任意选择 A，满足 $i \le A$，这样，第 $A, A+1, \ldots, A+k-1$ 天都可以用这张票进入体育馆。

q 次询问，每次给出 l, r，问如果第 l 天某人来到这个城市（也就是无法在第 l 天前买票），并且要在第 $l, l+1, \ldots, r$ 天进入体育馆，至少要花多少钱。

$$n, q \le 300000$$

提示：先形式化地写出问题。

单调栈

体育馆问题 CF1601E

$$\sum_{i\in\mathbb{N}}[l+ik\le r]\min_{l\le j\le l+ik}a_j$$

注意到最小值的计算除了第一天 a_l，都是 k 个 k 个一组的。同时，不妨设 l,r 模 k 同余。

$\min_{l\le j\le l+ik}a_j$ 可以写成 $\min(a_l,b_{l+k},\dots,b_{l+ik})$，其中 $b_i=\min_{i-k<j\le i}a_j$。可以用单调队列 $O(n)$ 求出 b。

思考：此时能否再简化问题？
提示：发现独立性。

单调栈

体育馆问题 CF1601E

$$\sum_{i \in \mathbb{N}} [l + ik \le r] \min_{l \le j \le l+ik} a_j$$

发现独立性。

注意到模 k 不同余的 a, b 对询问互不影响，所以可以分开处理每种模 k 的余数。

分开处理后，即为：给定一个序列，区间询问区间每个位置的 min(区间前缀最小值, v) 之和。
在单调栈上二分，找到 min 的分界点，前缀和回答。

单调栈

小结

单调栈的学习中，需要注意的要点有：

1. 理解单调栈维护的是什么：$a[1 \dots i]$ 的所有后缀最值。
2. 理解 l, r 数组的用途。当需要对所有区间的最值进行操作时，枚举最值 i 并考察 l_i, r_i 是常见的做法。
   特别地，理解区间最大值之和 l, r 需要一边取等一边不取等。
3. 理解单调栈上前缀和、二分是如何解决区间询问问题。

感兴趣的同学可以课后自学“笛卡尔树”。

序列分治梗概

分治是一个很大的话题，涵盖了广泛的算法设计思想。今天我们主要讲序列分治。

如果有一个问题需要对所有 $1 \leq l \leq r \leq n$ 的每一对 (l, r)，求它们整体的结果（例如区间最大值之和），就可以使用序列分治。

通常情况下，序列分治算法流程为：定义 Solve(L, R) 表示求出 $L \leq l \leq r \leq R$ 的整体结果。令 $mid = \frac{L+R}{2}$（取中点目的是减少递归层数），先求出 $L \leq l \leq mid, mid < r \leq R$ 的整体结果，再调用 Solve(L, mid) 和 Solve(mid+1, r)。

分治

序列分治和线段树

你也许已经发现，序列分治的划分方式和线段树类似。

思考：两者有没有其它共同之处？

(l, r) 的贡献被计算的 $[L, R]$ 其实就是在线段树上查询 $[l, r]$ 时，斜体部分同时满足的最浅结点！

```
int Sum(int p, int l, int r, int x, int y) {
    if (x <= l && r <= y) return sum[p];
    int mid = (l + r) >> 1, res = 0;
    if (x <= mid) res += Sum(p * 2, l, mid, x, y);
    if (mid < y) res += Sum(p * 2 + 1, mid + 1, r, x, y);
    return res;
}
```

分治

分治的目的

为什么分治能简化问题?

1. 原来的限制 $l < r$ 简化成了 $l \in A, r \in B$。
2. 对于与序列的区间 $[l, r]$ 相关的问题，左右两边可能具有独立性，且只需要在最后进行一次信息合并。相对地，线段树就需要 $\log n$ 次。
3. 对于与序列的区间 $[l, r]$ 的所有子区间 $[l', r']$ 相关的问题，将问题拆成了一个前缀、一个后缀和跨过的部分。前缀和后缀比区间简单，而跨过的就去掉了 $l' \leq r'$ 的限制。

逆序对 经典问题

求一个排列的逆序对数。

就是求有多少对 (i,j) 满足 $i<j$ 且 $a_i>a_j$。

分治后，问题变为：有多少对 $a_i\in S, a_j\in T$ 满足 $a_i>a_j$。

可以分别排序后再双指针。

你或许已经发现了，该过程和归并排序求逆序对的过程本质上是一样的。不过，与这里所讲相比，归并排序巧妙利用分治结构直接进行了排序，在朴素分治基础上优化了一个 log。

区间最大子段和 经典问题

给定一个序列，多次询问区间最大子段和。

分治，在 $l \in [L, mid], r \in [mid+1, R]$ 的时候处理询问 $[l, r]$。实现可以递归传一个 vector，其中包含所有还未解决且被当前区间包含的询问。这也是分治解决区间询问问题的一般方法。

对于询问 (l, r)，答案就是 $\max\{[L, mid]$后缀最大子段和$, [mid+1, R]$前缀最大子段和$, [L, mid]$最大后缀和 $+ [mid+1, R]$最大前缀和$\}$，这三项都可以在分支过程中扫描得出。

思考：时间复杂度？

$$O((n+q)\log n)$$

P6240

给定一个序列，每个元素是物品，具有体积 w_i 和价值 c_i。q 次询问，每次询问一个区间 $[l,r]$ 和背包大小 V，请你求出用这个背包去装区间内的物品（每个物品只能装一次），价值和最大是多少。

$$n \le 10^4, q \le 10^5, V \le 500$$

分治，在 $l \in [L, mid], r \in [mid+1, R]$ 的时候处理询问 $[l,r]$。具体如何处理？

对于询问 (l,r)，答案就是 $[l,r]$ 区间背包的一项，可以由左侧后缀背包和右侧前缀背包 $O(V)$ 合并而来。

分治

序列分治

P6240

给定一个序列，每个元素是物品，具有体积 w_i 和价值 c_i。q 次询问，每次询问一个区间 $[l,r]$ 和背包大小 V，请你求出用这个背包去装区间内的物品（每个物品只能装一次），价值和最大是多少。

$$n \leq 10^4, q \leq 10^5, V \leq 500$$

思考：

1. 上述算法的时间复杂度？

   $$O(nV\log n + qV + q\log n)$$

2. 本题中分治和线段树相比体现了什么特点？

   只需一次合并的一项，而线段树需 log 次合并的 $O(V)$ 项

Special Segments CF1156E

给定一个数组，问有多少个 (i,j) 满足 $a_i + a_j = \max\limits_{k\in[i,j]} a_k$。

$n \le 10^5$

先分治，假设 $i \in [l, mid], j \in (mid, r]$。

预处理左边后缀 max 数组 c 和右边前缀 max 数组 d，问题化为

$$a_i + a_j = \max(c_i, d_j)$$

注意：这一步又体现了分治的常见用法是化多信息合并为二信息合并。
固定 i，哪些 j 会满足要求？

Special Segments CF1156E

给定一个数组，问有多少个 (i,j) 满足 $a_i + a_j = \max\limits_{k\in[i,j]} a_k$。

$n \le 10^5$

$$a_i + a_j = \max(c_i, d_j)$$

固定 i，哪些 j 会满足要求？

若 $c_i \ge d_j$，就是 $a_j = c_i - a_i$。
若 $c_i < d_j$，就是 $a_i = d_j - a_j$。

$c_i \ge d_j, c_i < d_j$ 的限制把 j 分成了 $j \le p, j > p$ 两端，p 可以随 i 双指针。
问题转化为：前缀/后缀 $= x$ 的元素个数，可以双指针过程中维护 map。

分治

小结

前面的例题已经向我们展示了分治解决区间计数问题的范式：

1. 转为两个信息合并。
2. 对合并方式分类讨论，再分离变量，分别计算。

这里，不要惧怕分类讨论繁复，毕竟“讨论”之难，怎样也不如“思维”之难。是谓：

胆大心细不畏繁，勇刃难题换笑颜。

更多练习题：P4755

最大独立集 P7482

给定一个序列，求所有区间的最大独立集之和。

$$n \le 10^5$$

看题，便知可立刻写出最大独立集的二信息合并形式：

预处理出 $l0, l1, r0, r1$ 分别表示左侧不选/选（这里，为了避免额外的取 max 操作，可以认为 1 表示选/不选均可）mid 的最大独立集，右侧不选/选 $mid+1$ 的最大独立集，则

$$ans(i,j) = \max(l1_i + r0_j, l0_i + r1_j)$$

提示：分类讨论，分离变量。

最大独立集 P7482

给定一个序列，求所有区间的最大独立集之和。

$$n \le 10^5$$

$$ans(i,j) = \max(l1_i + r0_j, l0_i + r1_j)$$

枚举 i。

若 $l1_i + r0_j \ge l0_i + r1_j$，就是 $l1_i - l0_i \ge r1_j - r0_j$。要求所有这样的 j 的 $l1_i + r0_j$ 之和，也就是分别是个数、$r0$ 之和。提前把 j 按照 $r1 - r0$ 排序，二分出分界点，维护前缀和即可；另一侧同理。

时间复杂度 $O(n \log^2 n)$。

偏序问题

偏序问题的定义

偏序问题，是形态如下的问题：

偏序问题的一般形式

k 维空间里有 n 个点，第 i 个点的坐标为 $(x_{i,j})$ 其中 $1 \le j \le k$。

有 q 次询问，每次询问包含一个点 (y_j)，求出有多少个给出的点满足 $\forall 1 \le j \le k, x_{i,j} \le y_j$。

有时，上述问题也简称“k 维数点”。

偏序问题

偏序问题的定义

有时题目要求的不是点数，而是点的权值和/异或和或更为复杂的信息合并，维护方法没有本质区别。

需要注意，偏序问题

1. 是“数点”，不是数“点对”或其他。
2. 只能处理“每维分别小于（当然小于等于也可以）”的限制。

解决偏序问题的通法叫做 CDQ 分治，可以在 $k-1$ 个 log 的时间复杂度内解决 k 维偏序。

偏序问题

偏序问题的例子

在具体学习 CDQ 分治的流程之前，我们先看几个偏序问题的例子。事实上，在省选及以前的阶段，OI 中的数据结构问题有很大一部分都用到了偏序问题的思想，所以学习它是非常有用的。

1. 逆序对计数
   把 (i, a_i) 看成点，则每个 i 对应的 j 的数量是二维数点。
   注意：不是“逆序对”是二维数点，而是对于一个 i 数 j 是二维数点。
2. 矩形求和
   给出二维平面上的一些点，每次询问一个横平竖直的矩形，求出矩形内有几个点。
   拆分成前缀的和差。（由此也看出，如果不是前缀问题，则需要信息有可减性）

偏序问题

区间数颜色 P4113

给出一个序列，多次询问区间中有几种不同的 a_i。

$$n, q \le 10^6$$

记 lst_i 表示 a_i 上次（$j < i$）出现的位置。

则区间内 a_i 的种数就是 $j \in [l, r], lst_j < l$ 的个数，因为此时每种 a_i 仅在第一次出现时有贡献。

偏序问题

例子

带修数点

经典问题

平面上初始没有点，你需要支持动态加点、删点，同时求矩形内点数。

$$n, q \le 10^5$$

加入时间维 t，在点出现时视为点 (x, y, t)，权值为 1；消失时权值为 -1。

询问的时间为 t'，就是求 $t < t'$ 的所有点的权值和。

偏序问题

三角形求和 经典问题

给定平面上 n 个点，每次询问给出 u, v, w，求满足 $x + y \le u, x \ge v, y \ge w$ 的点的个数。

$$n, q \le 10^6$$

画图可知，可以拆分为三个区域的和、差，其中每个区域都只有两维限制。因此，本题实际上是二维偏序。

偏序问题

小结

前几个偏序问题的应用中，有几个需要注意的地方：

1. 对于“出现几次”或类似的问题，考虑 lst。
2. 带修问题就是加时间维。
3. 并不是把问题转化为偏序问题就万事大吉了，有时候如果发现的性质太少，偏序问题维数过高，反而可能阻碍问题的继续思考。

偏序问题

扫描线

二维偏序是最简单的偏序问题，通常采用扫描线来解决。

想象一根竖着的数轴扫过平面，在扫到 $x = i$ 时，数轴的 j 位置就存储了 (i, j) 位置的信息。

放在区间上，扫描线可以理解成：枚举 r，同时维护一个数据结构，在扫到 $r = r_0$ 时，数据结构的 l 位置存储 $[l, r_0]$ 的区间信息。

扫描线解决二维偏序的方法是：按 x 从小到大考虑点和询问，用数据结构维护 y 一维。可以结合平面上矩形数点来理解。

偏序问题

矩形面积并 经典问题

平面上有 n 个矩形，求其面积并。

$$n \le 10^5, x_i, y_i \le 10^9$$

首先离散化。注意：此处离散化需要保留权值。

对 x 一维扫描线，对 y 一维维护线段树。在扫到 $x = k$（实际上对应了离散化数组里的 $v_k \sim v_{k+1}$ 这一段）时，希望在线段树上 u 位置处维护出 (k,u) 的覆盖情况。

维护方法是：对于矩形 (x_0,y_0,x_1,y_1)，在 x_0 处给 $[y_0,y_1]$ 加一（注意离散化后是个左闭右开区间），在 x_1+1 处减一。询问即为 ≥ 1 的位置个数。

偏序问题

扫描线

矩形面积并 经典问题

平面上有 n 个矩形，求其面积并。

$$n \leq 3 \times 10^5$$

≥ 1 的位置个数可以转化为 $= 0$ 的位置个数。

此时有技巧：注意到覆盖次数总是 ≥ 0，所以只需维护最小值个数（由于离散化，其实维护的是最小值对应的区间长度和）。

偏序问题

连续段问题 CF526F

给定一个排列，问有几个区间满足 $r-l=\max_{i\in[l,r]} a_i-\min_{i\in[l,r]} a_i$。

$$n\le 3\times 10^5$$

枚举 r，计算有多少个 l 满足 $r-l=\max\limits_{i\in[l,r]} a_i-\min\limits_{i\in[l,r]} a_i$。我们希望在线段书上维护出答案。

提示：先不管计数问题，怎么维护 $\max\limits_{i\in[l,r]} a_i-\min\limits_{i\in[l,r]} a_i$？也就是，希望在 r 处时，线段树上 l 位置上恰好等于 $\max\limits_{i\in[l,r]} a_i-\min\limits_{i\in[l,r]} a_i$。

偏序问题

连续段问题 CF526F

给定一个排列，问有几个区间满足 $r-l=\max_{i\in[l,r]} a_i-\min_{i\in[l,r]} a_i$。

$$n\le 3\times 10^5$$

希望在 r 处时，线段树上 l 位置上恰好等于 $\max\limits_{i\in[l,r]} a_i-\min\limits_{i\in[l,r]} a_i$。

从 $\max\limits_{i\in[l,r]} a_i$ 到 $\max\limits_{i\in[l,r+1]} a_i$ 的变化不就是单调栈所求吗！在弹栈的时候顺便执行一下线段树区间加（区间为 $(st_{top-1},st_{top}]$）操作就行了。

同理，$\max\limits_{i\in[l,r]} a_i-\min\limits_{i\in[l,r]} a_i-(r-l)$ 这个整体也是好维护的，但如何计数？

偏序问题

连续段问题

CF526F

给定一个排列，问有几个区间满足 $r-l=\max_{i\in[l,r]} a_i-\min_{i\in[l,r]} a_i$。

$$n \le 3\times 10^5$$

提示："区间 0 的个数"难以维护，但回想之前所学，什么值的个数容易维护？类比"矩形面积并"问题。

注意到 $\max\limits_{i\in[l,r]} a_i-\min\limits_{i\in[l,r]} a_i-(r-l)\ge 0$，因此可以维护区间最小值个数。

偏序问题

CDQ 分治

CDQ 分治的目的是，将 k 维偏序问题转化为 $k-1$ 维偏序问题。当 $k=2$ 时，就可以用扫描线解决了。

对第一维进行类似序列分治的分治。不妨假设询问的是 $x_i < x, \ldots$ 的点数，则点 x_i 对于询问 x 的贡献在 $[x_i, x]$ 这个区间跨过 mid 时计算。

可以总结为：CDQ 分治特指考虑左侧对右侧的影响的分治。

在偏序问题上，分治第一维后，第一维的限制不复存在，也就把问题转为了总大小乘了 log 的 $k-1$ 维偏序问题。

偏序问题

讨论

CDQ 分治思想简单，但当维数 ≥ 4 时就又难写又慢了。各位在准备写高维偏序代码前，一定要多想能不能多发现一些能够降维的性质。（例如：若 a 有单调性，则 $i \leq p$ 和 $a_i \leq q$ 是同一维限制！）

如果信息不可减，偏序问题什么时候仍然可做？

至多只有一维不是前缀/后缀限制

偏序问题

如何正确地排序 P8253

有一个 $4 \times n$ 的数组 $a_{i,j}$，定义

$$f(i,j) = \min_k (a_{k,i} + a_{k,j}) + \max_k (a_{k,i} + a_{k,j})$$

求 $\sum_{i,j \in [1,n]} f(i,j)$。

$$n \leq 2 \times 10^5$$

分开处理 min 和 max。

讨论 min 和 max 分别在哪里取到，再分离变量，就是三维偏序问题。
需要注意相等时不要算重：可以认为 k 更小的值更小。

更多练习题：P4169, P3157, P2487

小结

我们刚刚学习了偏序问题的解法，特别地，学习了扫描线解决问题的思路。其中有以下要点值得注意：

1. 将“求 0 的个数”转为求“区间最小值个数”。
2. 对于有 max min 的式子，分类讨论在哪里取到再分离变量是常见处理方法。

综合运用

例子

天天爱打卡 P9871

小 T 跑步打卡。共有 n 天，每天可以跑步或不跑步。能量值初始为 0，若某天选择跑步，则能量值减少 d。不能连续超过 k 天跑步。

有 m 条奖励：若他在第 $l_i \sim r_i$ 天都选择跑步，会得到 v_i 能量值。

问 n 天后能量值最高是多少。

$$n \le 10^9, m \le 10^5$$

先对 l, r 离散化。思考：具体如何离散化？

综合运用

例子

天天爱打卡 P9871

选择跑步，则能量值减少 d。不能连续超过 k 天跑步。
有 m 条奖励：若他在第 $l_i \sim r_i$ 天都选择跑步，会得到 v_i 能量值。

以 l_i, r_i 划分数轴（均为闭区间），则跑步一定是一段一段跑。

设 dp_i 表示第 $1 \sim i$ 段的最优决策下能量值最大值，如何转移？

提示：枚举最后一段跑步的连续段。

若不跑，$dp_{i-1} \to dp_i$。
若跑，枚举 $j \le i$，$dp_{j-2} - d\left(s_i - s_{j-1}\right) + V(j,i) \to dp_i$。这里要求 $s_i - s_{j-1} \le k$，可以二分求出分界点。

综合运用

例子

天天爱打卡 P9871

选择跑步，则能量值减少 d。不能连续超过 k 天跑步。
有 m 条奖励：若他在第 $l_i \sim r_i$ 天都选择跑步，会得到 v_i 能量值。

若不跑，$dp_{i-1} \to dp_i$。
若跑，枚举 $j \le i$，$dp_{j-2} - d(s_i - s_{j-1}) + V(j, i) \to dp_i$。

用扫描线的思想，在 r_i 处给 $j \le l_i$ 的 j 加上 v_i 的贡献，则上述转移就是区间最大值。

时间复杂度 $O(m \log m)$。
更多类似题目：P2605, CF1889C2

综合运用

小结

时间所限，今天只为大家展示了一道数据结构用于优化具体算法的例子，也就是去年联赛的最后一题。

如果你会设计简单的序列 dp，掌握了扫描线的思想，也能正确处理离散化的细节，这题对你来说应当是容易的。

用数据结构优化其它算法时，以下法则应当牢记于心：

想清楚数据结构中存储的数据的含义。
时间？下标？存储内容？合并方式？
我想求的答案满不满足我想用的数据结构的要求？