根号算法

范一隆

天津大学

2023 年 7 月 18 日

目录

目录

摘要

本节主要内容

在竞赛中，如何高效的组织和处理大规模数据是一个重要问题。
在数据维护问题中根号算法也越来越常见，例如分块维护或者“莫队算法”。分块是一种适用性高，易于实现且思维难度低的方法，能够适用于解决一些信息不能快速合并的题目。
本文介绍了常见的分块算法，同时通过几道题目探索根号算法在非数据结构题目中的应用，通过这种思想获得更优的算法。

解决的问题

分块适用于以下情况

- 不常见的数据统计问题
- 能够应对强制在线问题
- 序列或树形结构
- 对复杂度要求较低的问题
- 特殊数据问题

目录

如何分块

分块方法

设数组长度为 n，我们对其分块，我们选取连续的 H 个为一块，若 H 能够整除 n，那么我们恰好分成 $\frac{n}{H}$ 块，分别编号 $[1, \frac{n}{H}]$，若不能整除，我们将最后 $n\%H$ 个划分为一块，编号为 $[1, \lfloor \frac{n}{H} \rfloor + 1]$。

变量表示

- H 表示每个块中有 H 个元素
- $block[i]$ 表示第 i 个位置所属的块的编号
- $block[n]$ 表示总共有多少个块
- $L[i] = (i-1) * H + 1$ 表示第 i 个块的左端点
- $R[i] = min(i * H, n)$ 表示第 i 个块的右端点

什么时候能够使用分块

以区间统计问题为例

1. 数据规模大，暴力方法复杂度为 $O(1) - O(n)$
2. 常规的 $O(logn) - O(logn)$ 的数据结构难以维护
3. 需要数据结构嵌套或者需要更深层次的挖掘信息
4. 考场之上要求程序编写难度较低
5. 其他情况

几个简单的引理

Lemma

分块的方法，可选取每 H 个为一块，最后一块不足 H 个也按照一块计算，块内第一个元素为关键点，则块的数目不超过 $\lfloor \frac{n}{H} \rfloor + 1$。

Lemma

查询 $[L, R]$ 的信息，完整的块不会超过 $\lfloor \frac{n}{H} \rfloor$，单独的点不会超过 $2 * H$。

Lemma

通常情况下，为了平衡两种操作的复杂度，H 取得 $\sqrt{n}$ 时达到最优复杂度，这也是分块算法也被称为根号算法原因。

目录

1. 直接分块

按照要求分块并维护信息

将序列直接进行划分，解决一些简单的修改，查询操作。序列分块是最简单最常见的例子。

例题 1 序列求和
n 个数，两种操作
1、修改 $a_x = y$
2、查询 $[l, r]$ 的区间和
例题 2 弹飞绵羊
n 个数，满足 $i < a_i \leq N + 1$，下面两种操作
1、修改一个数 a_i
2、查询从 x 开始跳,$x = a_x$，几次能够跳出序列

题目分析

例题 1 分析

我们记录每一个块内的总和是多少，查询时只需要查询完整的块和单独的点，复杂度 $O(n\sqrt{n})$

例题 2 分析

如果我们记录答案数组，那么每次修改操作最坏要跳 $O(n)$，查询操作 $O(1)$，如果我们 $O(1)$ 修改，那么每次查询最坏情况要跳 $O(n)$，所以我们考虑平衡两者的复杂度，每次修改往后面跳出当前块就停止，每次查询直接跳过完整的块，两者的负责度最坏情况都是 $O(\sqrt{n})$

2. 小区间维护数据结构

块内维护各种数据结构

在每个小块内进行一些数据结构的维护，或者是其他算法。

例题 3 数组变换

n 个数，m 次操作，查询 $[L, R]$ 区间严格小于 p 的数的个数 k，并在查询结束后将 $a[p]$ 修改为 k

题目分析

例题 3 分析

我们将序列分块，每个块内排序，查询时每个完整的块内二分查找 p，剩余的单点暴力枚举，修改时只需最多 $\sqrt{n}$ 次交换即可达到有序，复杂度 $O(n\sqrt{n}\log\sqrt{n})$

3. 分类维护与询问

按照数据大小进行分类

按照数据的大学分类讨论，将询问，修改等操作分成大于 H 的部分和不大于 H 的部分，结合两种暴力的方法，实现简单，思维复杂度低。
这类题目通常以两种形式出现：
第一种是出现次数大于 $\sqrt{n}$ 次的数的个数不会超过 $\sqrt{n}$
第二种是每次跳超过 $\sqrt{n}$ 距离跳的次数不会超过 $\sqrt{n}$

例题 4 序列求和 2
n 个数，两种操作
1、从第一个数起，每隔 D_i 个位置上增加 x_i
2、查询 $[l, r]$ 的区间和

题目分析

例题 4 分析

修改操作，根据 D_i 的大小划分，如果 D_i 大于 $\sqrt{n}$ 直接暴力修改，否则记录每个 D_i 的 ADD 操作的修改量的和，结合两种情况，修改复杂度为 $O(\sqrt{n})$。
查询操作，分块求和，对于所有小于 $\sqrt{n}$ 的 D_i，可以 $O(1)$ 算出它对 $[L, R]$ 区间的贡献，大于 $\sqrt{n}$ 的 D_i 使用例题 1 的方法，查询复杂度 $O(\sqrt(n))$。

4. 预处理

提前记录我们需要的信息

预处理是指维护一个数据集合时，在所有询问操作以前对数据进行一些处理往往记录许多附加信息，在询问时利用已处理的信息提高时间效率。

例题 5

n 个数，$a_i \le n$，多次询问 $[L, R]$ 中出现恰好 k 次的数的数目

Lemma

对于任意的两个区间 $[L1, R1], [L2, R2]$，如果记录了在区间 $[L1, R1]$ 中 $1 - n$ 之间每个数的出现次数以及出现次数为 $1 - n$ 的数的个数，那么可以在 $O(|L1 - L2| + |R1 - R2|)$ 的复杂度内将信息转化为 $[L2, R2]$ 这个区间的信息。

题目分析

例题 5 分析

对于出现次数超过 $\sqrt{n}$ 的数，我们预处理出这些数出现次数的前缀和，时空复杂度均为 $O(n\sqrt{n})$，查询时可以 $\sqrt{n}$ 判断这些数的出现次数是否等于 k。
对于小于 $\sqrt{n}$ 的询问，我们可以预处理出每个块出现次数为 $1-n$ 的数的个数和每个数的出现次数的前缀和，复杂度 $O(n\sqrt{n})$，那么我们可以 $O(1)$ 的得到完整的两个块之间的答案，再利用上页引理 $O(\sqrt{n})$ 时间得到查询区间的答案。
具体听课上讲解。

5. 离线统计

保存所有的查询

有些数据处理不要求对每次询问即时得到答案，而是一次性给大量询问，没有修改，我们记录下所有的查询，并对其进行一定排序，得到答案后统一输出，这就是离线算法，这类算法常见的代表就是莫队算法，莫队算法实际上是利用分块的特性进行的暴力统计，后面我们会进行讲解。

见例题 5

6. 升单项降整体

两者总体复杂度平衡

序列统计问题基本上都是修改与询问一起的，所以两者复杂度的平衡就很重要，如果修改的次数很多，那我们就要考虑适当提升查询复杂度的同时降低修改的复杂度，如果查询很多，就要降低查询的复杂度，以达到整体复杂度最优。该问题在莫队中比较常见。

例题 6

n 个数 a_i，还有 n 个函数 $f_i = \sum_{j=Li}^{Ri} a_j$，$m$ 次操作

1、修改 a_i

2、查询 f 数组的区间和

题目分析

例题 6 分析

预处理部分不做重点介绍，我们着重来看查询部分，将每次查询 $[L, R]$ 区间分成单点和完整的块，完整的块我们可以通过预处理 $O(1)$ 得到，但是每个单点 f 都是 a 数组的区间和，我们不能使用 log 的方法，所以我们使用 $O(1)$ 的前缀和，那么修改时，就是将 x 位置后面所有的都加上一个数，与 x 同一块的暴力修改，后面完整的块可以统一记录一下，复杂度变为 $O(\sqrt{n})$，整体的复杂度为 $O(n\sqrt{n})$。

7. 定期重建

达到某个操作次数后将状态更新

定期重建实际上就是对操作分块，每个操作块完成后就更新当前状态，在查询时根据上次更新的状态和当前块内修改对当前的影响计算答案。见例题 6

题目练习

练习 6.1
n 个数 a_i，m 次操作，每次给出两个整数 x, y，查询 $a[x+ky]$ 的和，其中 $k \in [0, \lfloor \frac{n-x}{y} \rfloor]$

练习 6.2
n 个数 a_i，m 次操作
1、修改 $[L, R]$ 区间，第 i 个数增加 $i - L + 1$
2、查询 $[L, R]$ 的区间和

练习 6.3
n 个数 $a_i \leq n$，m 次操作，每次都是 L, R, A, B 的形式，表示查询 $[L, R]$ 区间在 $[A, B]$ 范围内的书的个数。

练习 6.4
n 个数，m 次操作，两种操作，修改所有 k 的倍数的位置，或者区间求和。

目录

算法简介

算法简介

- 莫队是由莫涛归纳整理的一种区间操作算法，以简短的框架和优秀的复杂度闻名。
- 莫队真正的算法精髓在于复杂度的分析。
- 实际上是一个优化版本的暴力算法。
- 学习莫队算法的前提要了解序列分块。
- 了解在线、离线、强制在线算法

算法原理

形式

假设 $n = m$，那么对于序列上的区间询问问题，如果从 $[l, r]$ 的答案能够 $O(T)[l-1, r], [l+1, r], [l, r+1], [l, r-1]$（即与 $[l, r]$ 相邻的区间）的答案，那么可以在 $O(n\sqrt{n}T)$ 的复杂度内求出所有询问的答案。

解释

离线后排序，顺序处理每个询问，暴力从上一个区间的答案转移到下一个区间答案（一步一步移动即可）。

排序方法

对于区间 $[l, r]$, 以 l 所在块的编号为第一关键字，r 为第二关键字从小到大排序。

算法原理

复杂度分析

- 认为 n, m 同阶
- 排序复杂度
- 第一个询问的复杂度
- 块内的复杂度
- 不同块的复杂度

移动顺序

l,r 两个指针有四个移动方向，要保证 $--l$ 在 $r--$ 前，$++r$ 在 $l++$ 前

优化

按照左端点所在块奇偶分类排序

莫队算法

例题 7 数颜色

n 个数，$a_i \le n$，多次询问 $[L, R]$ 中有多少个不同的数

题目分析

例题 5 分析

裸的莫队，可以看一下模版

```
void add(int x){
    cnt[x]++;
    if(cnt[x]==1) ans++;
}
void del(int x){
    cnt[x]--;
    if(cnt[x]==0) ans--;
}
int l=1,r=0,ans=0;
for(int i=1;i<=n;i++){
    while(l>q[i].l) add(a[--l]);
    while(r<q[i].r) add(a[++r]);
    while(l<q[i].l) del(a[l++]);
    while(r>q[i].r) del(a[r--]);
    q[i].ans=ans;
}
```

莫队算法

例题 8 排队接水

n 个小朋友，每个小朋友都有一个接水需要的时间 a_i，每次询问一个区间 $[L, R]$，任意安排这个区间的小朋友的接水顺序，使得总等待值最小。$n \leq 200000$

题目分析

例题 5 分析

每个区间按照接水时间从小到大排序是最优的，每次转移相当于区间内插入一个数或者删除一个数。用树状数组维护信息，复杂度 $O(n\sqrt{n}\log n)$

带修改莫队

算法简介

普通莫队是不支持修改操作的，可以通过对每个询问加上时间属性，表示经过的修改次数。
即 $[l, r]$ 变为 $[l, r, time]$，转移变成了六个方向。

排序方法

按对于区间 $[l, r, time]$，我们以 l 所在块的编号为第一关键字，以 r 所在块的编号为第二关键字，以 $time$ 为第三关键字。

时间复杂度

块大小为 $n^{2/3}$，时间复杂度为 $O(n^{5/3})$。

带修改莫队

例题 9 数颜色 2

加入了修改操作，每次修改就是将序列中的某一位修改为某个值，每次查询仍然是查询一个区间内有多少个不同的数。

题目分析

例题 5 分析

普通莫队的做法见例题 7，现在来考虑修改：

单点修改，把某一位的数字修改掉。假如我们是从一个经历修改次数为 i 的询问转移到一个经历修改次数为 j 的询问上，且 i<j 的话，我们就需要把第 i+1 个到第 j 个修改强行加上。

假如 j<i 的话，则需要把第 i 个到第 j+1 个修改强行还原。

怎么强行加上一个修改呢？假设一个修改是修改第 pos 个位置上的颜色，原本 pos 上的颜色为 a，修改后颜色为 b，还假设当前莫队的区间扩展到了 [l,r]。

加上这个修改：我们首先判断 pos 是否在区间 [l,r] 内。如果是的话，我们等于是从区间中删掉颜色 a，加上颜色 b，并且当前颜色序列的第 pos 项的颜色改成 b。如果不在区间 [l,r] 内的话，我们就直接修改当前颜色序列的第 pos 项为 b。

还原这个修改：等于加上一个修改第 pos 项、把颜色 b 改成颜色 a 的修改。

其他莫队

例题 5 分析

- 树上莫队
- 树上带修改莫队
- 回滚莫队

目录

根号算法

题目解析

例题 10 不同的 $\lfloor \frac{n}{i} \rfloor$
例题 11 $\sum_{i=1}^{n} \lfloor \frac{n}{i} \rfloor$
例题 12 $\sum_{i=1}^{n} n\%i$
例题 13 $Ans = max\{gcd(i,j)|i \in [a,b], j \in [c,d]\}$
例题 14 给一个长度为 N($n \leq 300$) 的 01 字符串，再给一个正整数 M，每次操作可以将一个位置或者将一个长度为 M 的倍数的前缀取反，问最少多少次操作可以变成一个循环节为 M 的循环串。
例题 15 给出一个 N 个点 M 条边的无向图。初始时所有点为黑色，两种操作，将一个点反色，或者查询某个点相连了多少个黑点。

目录

分块习题

简单题目

[*cogs* 1689] 弹飞绵羊
[*bzoj* 3343] 教主的魔法

中等题目

[*cogs* 1695] 梦游仙境
[*bzoj* 3295] 动态逆序队

有思维含量的题目

[*bzoj* 2141] 排队
[*cogs* 2312] 简单的求和问题
[*bzoj* 2724] 蒲公英

莫队习题

序列维护类

[*bzoj* 2120] 数颜色
[*bzoj* 3781] 小 B 的询问
[*bzoj* 2038] 小 Z 的袜子
[*bzoj* 3809]Gty 的二逼妹子序列
[*bzoj* 3236] 作业
[*bzoj* 3289]Mato 的文件管理

与其他算法结合类

[*bzoj* 3585]mex
[*bzoj* 4540] 序列
[*hdu* 4638]Group
[51*nod*1592] 数列积

目录

Thank you! Any Question?