

对一类「求区间内不同元素的个数」问题的分析

上海市育才中学 丁梓洋

摘 要

一类「求区间内不同元素的个数」问题是大数据结构和离线算法入门的经典例题。

近年来，大数据结构题在信息学奥赛中总以压轴题出现。本文通过介绍一类「求区间内不同元素的个数」的问题，分析比较了几种不同的解法，希望能够帮助读者更好地理解此类问题的解决方法。

关键词 莫队算法 CDQ 分治 树状数组 线段树 ZKW 线段树 树套树 珂
朵莉树

目 录

第一章 区间询问	1
1.1 题目大意	1
1.2 分析	1
1.3 解法	1
1.3.1 树状数组 / 线段树	1
1.3.2 分块	2
1.3.3 分治	2
1.3.4 莫队算法	2
1.3.5 主席树	2
1.3.6 扫描线	2
1.3.7 归并树	2
第二章 区间询问、单点修改	3
2.1 带修莫队	3
2.2 CDQ 分治	3
2.3 分块	3
2.4 ZKW 线段树	3
2.5 树套树	3
2.6 珂朵莉树	3
第三章 区间询问、区间修改	3
参考文献	5
致 谢	5
附 录	6

第一章 区间询问

1.1 题目大意

给定一个长为 N ($1 \leq N \leq 10^5$) 的序列 $\{a_i\}$ ($1 \leq a_i \leq 10^6$), 有如下指令:

- $Q\ l\ r$: 询问区间 $[l, r]$ 内有多少种不同的数。

现有 M ($1 \leq M \leq 10^5$) 条指令。要求对每条询问指令输出其答案。内存限制 512 MB, 时间限制 2.5 s。¹

1.2 分析

这个问题属于典型的区间查询问题, 要求高效地处理大量查询。由于序列长度和查询数量都非常大, 若采用暴力解法, 直接遍历区间计算答案将会导致时间复杂度过高, 超出了时间限制。

因此, 采用高效的数据结构和离线的思想来优化查询过程是必要的。常见的方案包括树状数组、线段树、莫队算法等。这些数据结构都可以高效地支持区间查询和单点更新, 因此它们成为了解决这类问题的基础。

1.3 解法

1.3.1 树状数组 / 线段树

首先, 我们需要预处理序列中每个元素上一次出现的位置。记序列第 i 位的值为 a_i , a_i 上一次在序列中出现的位置为 p_i 。这个预处理可以在读入序列时完成, 时间复杂度为 $O(n)$ 。

我们可以考虑将询问离线。具体来说, 将每个查询区间按照其右端点 r 升序排序。这样的离线可以在处理查询时, 利用之前处理的结果, 避免重复计算。记排序后第 i 个询问为 $[l_i, r_i]$ 。

我们可以使用树状数组来维护每个位置的状态。树状数组可以实现在 $O(\log N)$ 的时间复杂度内完成单点更新和区间查询的操作。

¹来源: 洛谷 P1972 [SDOI2009] HH 的项链, 有修改

每次根据 r 的升序计算询问区间的同时，更新树状数组。假设上一个处理的区间为 $[l_i, r_i]$ ，则现在处理的区间为 $[l_{i+1}, r_{i+1}]$ ，我们需要更新树状数组 $[c_{r_i}, c_{r_{i+1}}]$ 区间的每个值。

对于相同的值，我们只关心这个值在区间中出现的最右一个。如果 a_i 在 $[1, i)$ 之间没有出现过（即 $p_i = 0$ ），则将 c_i 记为 1；如果 a_i 在 $[1, i)$ 之间出现过（即 $p_i \neq 0$ ），则将 c_i 记为 1 的同时将 c_{p_i} 记为 0（在维护树状数组时，通过将 c_i 加 1 或减 1 的方式实现）。因此，我们就避免了重复计算同一个数值。 $[l_i, r_i]$ 区间内的不同数的个数即为 $\sum_{j=l_i}^{r_i} c_j$ （树状数组通过 $\text{query}(r_i) - \text{query}(l_i - 1)$ 实现）。

除了树状数组，我们还可以使用线段树来实现，实现方法与树状数组类似。线段树的每个节点维护区间内的不同数的个数。在更新和查询线段树时，我们可以通过递归地更新或查询左右子树来实现。其常数时间复杂度较树状数组更大，且代码实现较为复杂。

该算法的时间复杂度为 $O((N + M) \log N)$ ，其中 N 为序列长度， M 为询问区间的数量。预处理 $\{p_i\}$ 的时间复杂度为 $O(n)$ 。将查询按照右端点 r 排序的时间复杂度为 $O(M \log M)$ 。对于所有查询，更新树状数组或线段树的时间复杂度为 $O((N + M) \log N)$ 。

1.3.2 分块

<https://www.luogu.com.cn/article/4vp3j5v4>

1.3.3 分治

<https://www.luogu.com.cn/article/9ajn0u1i>

1.3.4 莫队算法

<https://www.luogu.com.cn/article/306d31qt>

1.3.5 主席树

TODO

1.3.6 扫描线

<https://www.luogu.com.cn/article/j7n2fv29>

1.3.7 归并树

<https://www.luogu.com.cn/article/rzuy3es9>

第二章 区间询问、单点修改

给定一个长为 N ($1 \leq N \leq 10^5$) 的序列 $\{a_i\}$ ($1 \leq a_i \leq 10^6$), 有如下两种指令:

- $Q\ l\ r$: 询问区间 $[l, r]$ 内有多少种不同的数。
- $R\ i\ x$: 将 a_i 替换为 x 。

现有 M ($1 \leq M \leq 10^5$) 条指令。要求对每条询问指令输出其答案。内存限制 512 MB, 时间限制 2.5 s。²

2.1 带修莫队

<https://www.luogu.com.cn/article/tkzm8297>

2.2 CDQ 分治

<https://www.luogu.com.cn/article/hzoyp8rd>

2.3 分块

<https://www.luogu.com.cn/article/6o26r3fu>

2.4 ZKW 线段树

<https://www.luogu.com.cn/article/um3phhxf>

2.5 树套树

<https://www.luogu.com.cn/article/fp3jadsk>

2.6 珂朵莉树

<https://oi-wiki.org/misc/odt/>

第三章 区间询问、区间修改

给定一个长为 N ($1 \leq N \leq 10^5$) 的序列 $\{a_i\}$ ($1 \leq a_i \leq 10^9$), 有如下两种指令:

²来源: 洛谷 P1903 [国家集训队] 数颜色 / 维护队列, 有修改

- $Q\ l\ r$: 询问区间 $[l, r]$ 内有多少种不同的数。
- $R\ l\ r\ x$: 将序列中 $[l, r]$ 内的所有数都替换为 x 。

现有 M ($1 \leq M \leq 10^5$) 条指令。要求对每条询问指令输出其答案。内存限制 64 MB, 时间限制 1.5 s。³

³来源: 洛谷 P4690 [YNOI2016] 镜子里的昆虫, 有修改

参考文献

致 谢

- 感谢「沪疆教育信息化 算法小论文」提供的学习和交流的平台。
- 感谢张卫国老师、诸峰老师对我的指导和帮助。
- 感谢洛谷平台提供的题目和测试。

附 录

附录 1

这是一个附录内容，学校规定附录的二级标题得是“附录”二字后接阿拉伯数字。

但是 Typst 的中文与英文和数字之间的空格并没有像 LaTeX 那样自动空出，所以就需自己手打了。