JZPLCM命题报告

南京师范大学附属中学　顾昱洲

Abstract

区间询问问题是信息学竞赛中经常出现的一类问题，数论也是近年NOI系列竞赛中的一个热点。作者将区间询问问题与数论结合，出了一道颇有新意的题目。本文详细介绍了本题的解法、出题思路以及其他相关内容。

Keywords: 最小公倍数, 区间询问问题

Contents

1 题目大意

有一正整数序列$a_1, a_2, \dots, a_n$。有q次询问，每次询问一个区间内所有数的lcm(最小公倍数)。由于答案可能很大，输出答案模10^9+7。

$n, q \le 10^5, 1 \le a_i \le 10^9$

部分分的数据范围参考题目描述。

2 初步分析

首先进行一些约定。

显而易见的，我们要对输入中的所有数分解质因数。由于数的范围非常小，所以我们直接暴力枚举质因数分解即可(比pollard's rho快)。这一部分的复杂度是$O(n\sqrt{C}/\log C)$，其中C为a_i的上界。以下复杂度分析中不包含分解质因数的复杂度。

一个数的不同的质因子个数为$O(\log C)$；一个数的质因子的幂的和为$O(\log C)$。

我们考虑lcm的性质。假设我们要求一些数的lcm。那么，对于每个质数p，要求的数中p的最大的幂为q，那么对答案的贡献即为p^q。答案即为所有贡献的积。

由此，我们可以得到算法一：

2.1 算法一

对于每个询问依次处理，每次对每个数的每个质因子，更新其贡献。

时间复杂度$O(qn\log n)$。期望得分20。

对算法一进行简单优化，可以得到算法二：

2.2 算法二

对于每个询问依次处理。枚举每个质数p，这样我们只要求一段区间内幂的最大值。这个是经典的RMQ问题。考虑到本题的性质，可以使用ST表。

时间复杂度$O(wn\log n + wq)$，其中w为需要枚举的质数的个数。期望得分40(结合算法一)。

3 通用算法

本题是一个区间查询问题。我们常常见到看上去较为困难的区间查询问题。例

如：

- 无修改区间众数查询
- 莫涛《小Z的袜子》
- 徐捷《图中图》
- Yandex Algorithm 2011 Round 2 D Powerful Array (CF86D)

对于这些问题，我们有一些通用的算法。

3.1 算法三

将每个询问$[l,r]$看成平面上的点(l,r)。对于所有询问排序，使得相邻两个询问的曼哈顿距离之和为$O((n+q)\sqrt{n})$。

一种简单的方法是，将所有询问以$\lfloor (r-l)/\sqrt{n} \rfloor$为第一关键字，$l$为第二关键字排序。

然后，我们按照排序后的顺序对每个询问处理，在进行每次询问的同时充分利用上一次询问的结果(离线)。

可以发现，我们每次只要对一个集合S(初始为$\varnothing$)支持如下操作：

- 插入一个数
- 删除一个数
- 维护S中所有数的答案

本题中，即为维护每个质数的出现的最大的幂，由于有删除，需要使用平衡树。

时间复杂度$O((n+q)\sqrt{n}\log^2 n)$。由于常数较大，期望得分60(结合算法二)。

3.2 算法四

在序列上，等距离取L个关键点。我们先通过算法一中的方法预处理出关键点两两之间的答案。这一部分的时间复杂度是$O(Ln\log n)$。

然后，对于每个询问$[l,r]$。我们取l右边最左的关键点l'和r左边最右的关键点r'。l'到r'的答案已经预处理出。现在，只要对$[l,l')$和$(r',r]$进行一遍算法一中的方法即可。注意这里需要询问某质数在一段区间内的最大的幂。

这一部分的时间复杂度可以做到$O(qn\log n/L)$。可以发现，L取$\sqrt{q}$时达到最优。

总时间复杂度$n\sqrt{q}\log n$(简单一些的做法，可以做到$O(n\sqrt{q}\log^{1.5} n)$)。

这个方法是在线的，且时间复杂度和常数均较算法三小，期望得分$60\sim 80$(结合算法二)。

4 进一步分析

我们考虑算法三和算法四的劣势。它们的劣势在于，它们能解决的东西太多

了(能够自然地修改以解决3中的问题)。算法过于一般化，导致对于特殊问题难以做到更好。

我们考虑解决问题的两种思路：

(1) 原问题→一般化(强化)的问题→解决一般化的问题→解决原问题

(2) 原问题→特殊化(弱化)的问题→解决特殊化的问题→推广到原问题

(需要注意的是，这里的特殊化、一般化与“转化”的意义不大一样)

之前的算法一二过于简单；算法三四采用了思路(1)，得到了一个可推广性很强，但是不够优秀的算法。

因此，我们可以尝试思路(2)：先解决一个与本问题类似，但是较简单的问题，然后试图推广到原问题。

4.1 JZPGCD

一个简单的想法是把本题中的lcm改为gcd。这个时候的问题是非常容易解决的。只要采用线段树即可。过于简单，不详细叙述。

然而，这个方法无法自然拓展到本题。因为子区间的gcd的结果很容易合并(直接求gcd即可)，然而子区间的lcm很难合并(因为要对答案取模)。

因此这样的弱化是不行的。

4.2 JZPSQF

如果输入中的a_i都是无平方因子数(Square-Free Number)，应该如何解决？

注意到，此时每个质数p的最大的幂只有0和1两种情况，也就是出现和不出现。这样，我们把每个数拆分成它的所有因子，问题转化成了某区间内所有出现的数的积(所有不同数的积)。

这是一个经典问题(经典问题的解法见4.4)，可以在理想的时间复杂度内解决。

这个方法可以自然地拓展到原问题，由此我们得到算法五：

4.3 算法五

同样的，拆分每个数。当前拆分x，如果其有因子p^q(p为质数，q为满足$p^q|x$的最大的q)，那么拆分出q个数$p^1, p^2, \ldots, p^q$，权值均为p。这样，问题转化成了某区间内不同的数对应的权值的积。

证明 对于每个质数p，假设它在询问的区间内出现的最大的幂为q，那么在该区间内与p有关的数是$p^1, p^2, \ldots, p^q$，权值均为p。这样，p对答案的贡献即为p^q，与lcm的性

质符合。

这依然是一个经典问题(经典问题的解法见4.4)。可以在$O(n\log^2 n+q\log n)$的时间复杂度内解决。期望得分100。

4.4 经典问题

4.4.1 题目大意

有一正整数序列$a_1,a_2,\ldots,a_n$，对应的权值分别为$b_1,b_2,\ldots,b_n$。且满足$a_i=a_j\Rightarrow b_i=b_j$(答案的唯一性)。

有q次询问，每次询问一段区间内不同的数对应的权值的积，答案模10^9+7。

4.4.2 算法分析

对于每个i，预处理出c_i，为满足$a_j=a_i$且$j<i$的最大的j(若不存在则为0)。

那么，求一个区间$[l,r]$内不同的数对应的权值的积，也就是$[l,r]$内$c_i<l$的b_i的积。

无修改离线做法 首先，把询问的区间$[l,r]$变成$[1,r]$减$[1,l-1]$，由于10^9+7是质数，且大于a_i的最小上界，因此这样做是可行的。

然后，按照右端点顺序依次处理每个询问(这时所有询问的左端点都是1)，这样，我们只要维护一个数据结构，支持：

- 插入一个数对(c_i,b_i)
- 询问所有$c_i<x$的数对b_i的积

显然树状数组可以支持这些操作(且常数非常小)。

时间复杂度$O((n+q)\log n)$，空间复杂度$O(n)$。

无修改在线做法一 使用线段树套有序数组。线段树的每个端点存储这个区间内所有(c_i,b_i)对按c_i排序的结果，以及b_i的前缀积。

时间复杂度$O(n\log n+q\log^2 n)$，空间复杂度$O(n\log n)$。

无修改在线做法二 采用和无修改离线做法类似的思路。但是此时我们要求在线，只能先预处理出每个时刻的数据结构。然后进行历史版本查询。这个可以使用函数式线段树解决。

时间复杂度$O(n\log n+q\log n)$，空间复杂度$O(n\log n)$。

有修改做法 如果有修改，可以采用无修改在线做法一中的思路，采用线段树套平衡树。

时间复杂度$O(n\log n+q\log^2 n)$，空间复杂度$O(n\log n)$。

5 总结与其他

5.1 思维方法

从解决本题的过程中，我们可以知道解决问题的一种思维方法。在其他方法无效时，适当(非平凡地)退化问题，解决退化问题，将退化的解法进化，得到原问题的解法。通过合理退让解决问题的方法，看似走了弯路，实际上给我们的思路带来了极大便利。

5.2 数据设计

本题的数据范围在题目描述中已经给出，这里不重复。主要阐述生成数据的方法。

本题的数据生成是较为容易的。

对于所有a_i均为不超过100的质数的积的数据，我采用的方法是：首先给初始值$x=1$，然后每次在$1\ldots 100$内随机选择一个质数p，若$xp\le 10^9$，那么将x更新为xp，否则结束。

对于其他数据，输入中的所有数(除了n和q)，均是随机生成的。随机数据中往往会有各种情况出现，因此非完美算法一般不能得到高分。

我曾经尝试过一些其他的构造方法，例如大部分a_i都是大质数或大部分a_i都是大质数的积等，然而除了使得分解质因数一步大大减慢之外，没有其他效果。因此没有采用。

算法五是可以支持在线或有修改的，但是那样纯粹只是增加代码复杂度，与NOI考察选手的思维水平的初衷相违背。因此本题是可离线无修改的。

5.3 得分估计

本题大致是NOI中难题水平。

本题的区分度较好，通过对数据范围的调控以及特殊类型数据的设定，不同水平的选手均能得到相应的分数。

如果作为NOI的试题，估计得分分布如下：

- 绝大部分选手能获得至少20分；
- 四分之一选手能够获得$40\sim 50$分；

- 十几名选手能够获得$60 \sim 80$分；
- $0 \sim 3$名选手能够获得$90 \sim 100$分。

这四类得分分别对应算法一、算法二、算法三四、算法五。

5.4 总结

本题是一道区间询问问题。区间询问问题在OI中经常出现，解法几乎已经格式化了(算法三四)。

在这种情况下，本题仍是一道有新意的问题。适当与数学结合，只需要基础的数学知识，不纠结于推导各种公式。最终得到的算法简洁优美，又不过于简单。

得到最终算法的思维方法也是在解决其他问题时值得借鉴的。