《支柱》解题报告

杭州学军中学 金策

1 试题来源

PA 2013 Round 6 Filary

提交地址：http://main.edu.pl/pl/archive/pa/2013/fil

BZOJ：http://www.lydsy.com/JudgeOnline/problem.php?id=3837

2 试题大意

给定n个正整数$w_1, w_2, \cdots w_n$，从中挑出k个数，满足：存在某一个$m \geq 2$，使得这k个数模m的余数相等。

求出k的最大值，并求出此时的m。

如果有多组解使得k最大，你要在此基础上求出m的最大值（保证这个值存在）。

数据规模：$2 \leq n \leq 10^5, 1 \leq w_i \leq 10^7$。

3 算法介绍

3.1 发现性质

稍微探索一下，可以发现这样的性质：最优解中$k \geq \left\lceil \frac{n}{2} \right\rceil$。

证明. 取$m = 2$。按奇偶性将w_i分成两类，其中必有一类的元素数量不小于$\left\lceil \frac{n}{2} \right\rceil$，取出这一类所有元素，即可得到一组合法解。 □

这一性质告诉我们，如果任意选取一个w_i，则它被包含在最优解中的概率至少是$\frac{1}{2}$。

如果我们能设计一个算法，对于给定的w_i，求出包含它的最优解。那么随机选取w_i计算并重复k次，答案错误的概率不会超过2^{-k}。取$k=10$左右就已足够可靠，可以AC。

3.2 包含给定值的最优解

设必须被包含的w_i的值为z。

对于剩下的某个w_i，若满足$w_i \equiv z \pmod m$，则有$m \big| w_i - z$。令$x_i = |w_i - z|$。对于正整数m，设集合$S_m = \left\{i : m \big| x_i\right\}$。则问题转化为：寻找一个$m$，使得$k = |S_m|$ 最大；并在此基础上使m尽量大。

3.2.1 k的最大值

我们先不考虑第二问m的最大值。这时，为了使k尽量大，不妨设选取的m是质数；否则，取m的一个质因子，得到的k不会更小。

将所有x_i质因数分解。对于所有涉及到的质数p求出对应的$|S_p|$。则$\max |S_p|$即为第一问答案。

为了高效分解质因数，只需先用线性筛法预处理出10^7范围内每个数的最小质因子。

3.2.2 m的最大值

现在若考虑m的最大值，则选取的m不一定是质数。

设最优解中m的质因数分解为$m = \prod_j p_j^{\alpha_j}$。根据定义，显然有$S_m \subseteq S_{p_j}$。又由于$m$是最优解，有$|S_m| \ge \left|S_{p_j}\right|$。从而，对于所有$j$，有$S_m = S_{p_j}$。

因此，我们可以把S_p相同的所有p都合并起来。判断两个集合S是否相等可以用比较哈希值的方法。

合并时还需要考虑的是每个质因子的次数α_j。这也很简单，只要取$\alpha = \min_i \max \left\{d : p^d \big| x_i\right\}$即可。这样我们就可以求得最大的$m$值。

3.3 时空复杂度

记$W = \max w_i$。线性筛预处理时间$O(W)$，分解质因数的时间为$O(W)$。从而总时间为$O(W + kn \log W)$，其中k是随机次数。

空间复杂度为$O(n+W)$。

4 总结

本题利用传统的方法很难得到足够优秀的算法。但经过思考分析后，我们可以发现关于最优解的一个性质，并利用它得到一个随机化的算法。