

新年的贺电 解题报告

杭州学军中学 吉如一

1 试题来源

出题人：杜瑜皓。题目可以在这里找到：<http://uoj.ac/problem/178>

2 试题大意

给出一个函数，这个函数的定义域是 $[0, 2^{32})$ 中的1024个不同的整数，它的值域是 $[0, 1024)$ 中的整数。同时有若干组询问，每次给出一个函数定义域内的数，询问它的函数值。

你需要写一个程序支持编码和解码的功能。

编码部分：输入第一行是`encode`，接下来1024行每行两个整数描述这个函数。你需要输出一个01串。

解码部分：输入第一行是`decode`，解析来一行是你的程序输出的01串。接着描述了若干组询问，你需要对每一个询问输出函数值。

给分方式：如果你的解码部分回答不正确或者程序没有正常运行，得0分。否则设你的01串长度是 n ，你可以获得如下分数：

得分	条件	得分	条件
1	$n \leq 10^5$	6	$n \leq 15000$
2	$n \leq 43008$	7	$n \leq 14000$
3	$n \leq 40000$	8	$n \leq 13000$
4	$n \leq 30000$	9	$n \leq 12750$
5	$n \leq 20000$	10	$n \leq 12500$

如果你的输出满足多个条件，取得分最高的。

一共有两类数据，其中前30分满足定义域中所有数都在 $[0, 1024)$ 中，后70分

没有特殊限制。

3 算法介绍

3.1 算法一

针对前30分数据，我们发现定义域就是 $[0, 1024)$ 中的所有整数，所以我们可以依次发0到1023的所有函数值。这样需要 1024×10 的长度的01串。可以获得前3个点的满分。

3.2 算法二

最简单的想法，在编码时直接把读入的两个数给传过去！即先是一个32位的值，然后是一个10位的函数值，接着再是一个32位的值...解码的时候只需要按照这个格式就能得到输入了。

这样我们需要 $1024 \times (10 + 32)$ 的长度的01串，每一个点可以获得2分。

3.3 算法三

为了用更短的长度来编码，因为键值使用的长度不多，所以我们来考虑减少传输键时使用的长度。

可以采用哈希的方法，如果能把键从 $[0, 2^{32})$ 范围哈希到 $[0, 1024)$ ，我们就可以用算法一的方法直接传输辣。

我们可以多次随机哈希函数，然后再01串的最开头加上随机次数。这样只要在编码的时候和解码的时候采用同样的随机方法，我们在解码的时候就能得到这个哈希函数了。接着只需要把所有询问的值求一下哈希值，就能够回答询问了。

看起来已经解决了这个问题？然而我们要找到这样的把 n 个数映射成 $[0, n)$ 的排列的哈希函数，期望随机次数是 $\frac{n^n}{n!}$ 。由斯特林公式 $n! \approx \sqrt{2\pi n}(\frac{n}{e})^n$ ，所以随机次数是 $O(\frac{e^n}{\sqrt{n}})$ 的，显然无法在时间限制内求得。

期望得分0分。

3.4 算法四

算法三有着很好的提示作用，我们可以使用其他的哈希方法，比如每一次把 n 个键映射到 $[0, 1)$ ，使得两个值的数目是相等的，这样就可以进行分治辣。

随机找到这样的函数的期望次数是 $2^n / \binom{n}{n/2}$ ，带入斯特林公式后化简就可以得到期望次数是 $O(\sqrt{n})$ 的，这是可以接受的。

为了解码，我们需要约定传随机次数使用的字节数，因为要求做到100%的正确率，所以我们必须要用足够长位数来传次数。可以对每一层约定一个足够大的常数，然后输出长度等于当前层的常数的01串来表示随机次数。

这样我们传输随机次数所需要的01串长度 $L(n) = L(\frac{n}{2}) + \log_2(\sqrt{n}) = O(n)$ ，但是因为常数比较大，所以只能获得50分到80分。而时间复杂度等于 $T(n) = 2T(\frac{n}{2}) + O(n\sqrt{n}) = n\sqrt{n}$ ，在时间限制内是可以运行完毕的。

3.5 算法五

可以发现算法四中传输随机次数时位数有很大的浪费，因为在大部分情况下随机次数都用不到那么多位。

所以我们可以使用另外一种方法来传输随机次数：每一层设定一个常数 l ，然后如果次数的位数不超过 l 位，就输出一个二进制位0然后再输出 l 位二进制数表示随机次数；否则如果超出了 x 位，我们先输出 x 个1，之后跟上一个0，最后再输出随机次数的二进制表示。这样就可以愉快的传输不定长的整数辣。

用这种方法我们就可以减少传输时串长的常数，可以获得更多的分数，但是离满分依然还有一定的距离。

3.6 算法六

可以发现在分治若干层之后，剩下的数已经非常少了，我们可以直接使用算法三方法来直接找到一个映射到排列的哈希函数，这样就可以减少传输的随机次数的个数了。

在实际写程序的时候可以在只剩下8个数的时候，使用算法三的方法（如果阈值取16就超时辣），然后再卡卡常数调调参数就能获得满分啦。

3.7 数据构造方法

因为询问可以直接询问每一个键的键值，所以随机数据的强度还是挺高的，直接一波大随机就好啦。不过应该还是要让不同的键值的数目比较多，同时又要存在多个键的键值相同的情况。

3.8 总结

这是一道很新颖的题目，在国内的OI比赛中似乎还没有这种编码解码的题目（好像只见过一道IOI题目是考察这方面的）。

使用哈希函数来进行对键进行哈希从而缩短传输长度，再传输的时候传输哈希函数的随机次数而不是哈希函数本身，使用前导0和1来传输长度不确定的整数，这些思想都是比较新颖而且也不是特别容易想到的，这些都提高了这个问题的难度。

同时为了选取合适的哈希方法，需要选手对自己的哈希函数进行分析——这相当于某种程度上的构造题——在一定程度上考察了数学知识，综合考察了选手的实力，从这些方面来讲感觉这是一道非常好的题目。