

数位计数问题的解法研究

北京市清华附中
高逸涵

引言

- 数位计数问题
 - 主要与数的各位数字构成有关
 - 统计一段连续区间内的数的性质
 - 完全模拟题目描述会严重超时

引言

- 此类问题的一般性解法：
 - 将整个区间划分为若干子段
 - 对于每个子段，通过子段性质直接求解
 - 合并各子段结果，得到总结果
- 以上为解决此类问题的总原则，接下来我们通过两道例题说明如何利用上述原则解决具体问题。

例题 1 : The Sum (SPOJ KPSUM)

- 将 $1 \sim N$ 内所有数按照从小到大顺序从左到右依次写下，然后在每一数位之前依次插入加号和减号（循环），求结果。
- 数据范围： $1 \leq N \leq 10^{15}$
- 举例： $N=11$ 时，答案为 $+1-2+3-4+5-6+7-8+9-1+0-1+1=4$

例题 1 : The Sum

- 显然直接模拟题目叙述并不是一个可行的策略，需要找到一种高效的算法。
- 因为加减符号的改变与数字个数相关，因此为了让规律更加明显，我们尽量将 $1 \sim N$ 划分为若干段区间，使得每个区间内的数的数字个数相同。

例题 1 : The Sum

- 按照上述原则将 $[1, N]$ 划分为若干子区间：
（这里以 $N=123456$ 为例）
- $[1, 9] \cup [10, 99] \cup [100, 999] \cup [1000, 9999] \cup [10000, 99999] \cup [100000, 123456]$

例题 1 : The Sum

- 那么，原问题转化为一个新问题：询问 $[A,B]$ 的结果，其中 **A** 和 **B** 包含相同的数字个数。
- 根据数字个数的奇偶性，这里分为两种情况讨论。

例题 1 : The Sum

- 数字个数为奇数的情况：（这里以 [10000,56789] 为例进行研究）

- $+1-0+0-0+0$

- $-1+0-0+0-1$

- $+1-0+0-0+2$

- $-1+0-0+0-3$

- $\dots\dots\dots$

- $-5+6-7+8-9$

可以看到，相邻两项基本都互相抵消了，只有个位相差
1

例题 1 : The Sum

- 数字个数为偶数的情况：（这里以 [100000,456789] 为例进行研究）

- $+1-0+0-0+0-0$

- $+1-0+0-0+0-1$

- $+1-0+0-0+0-2$

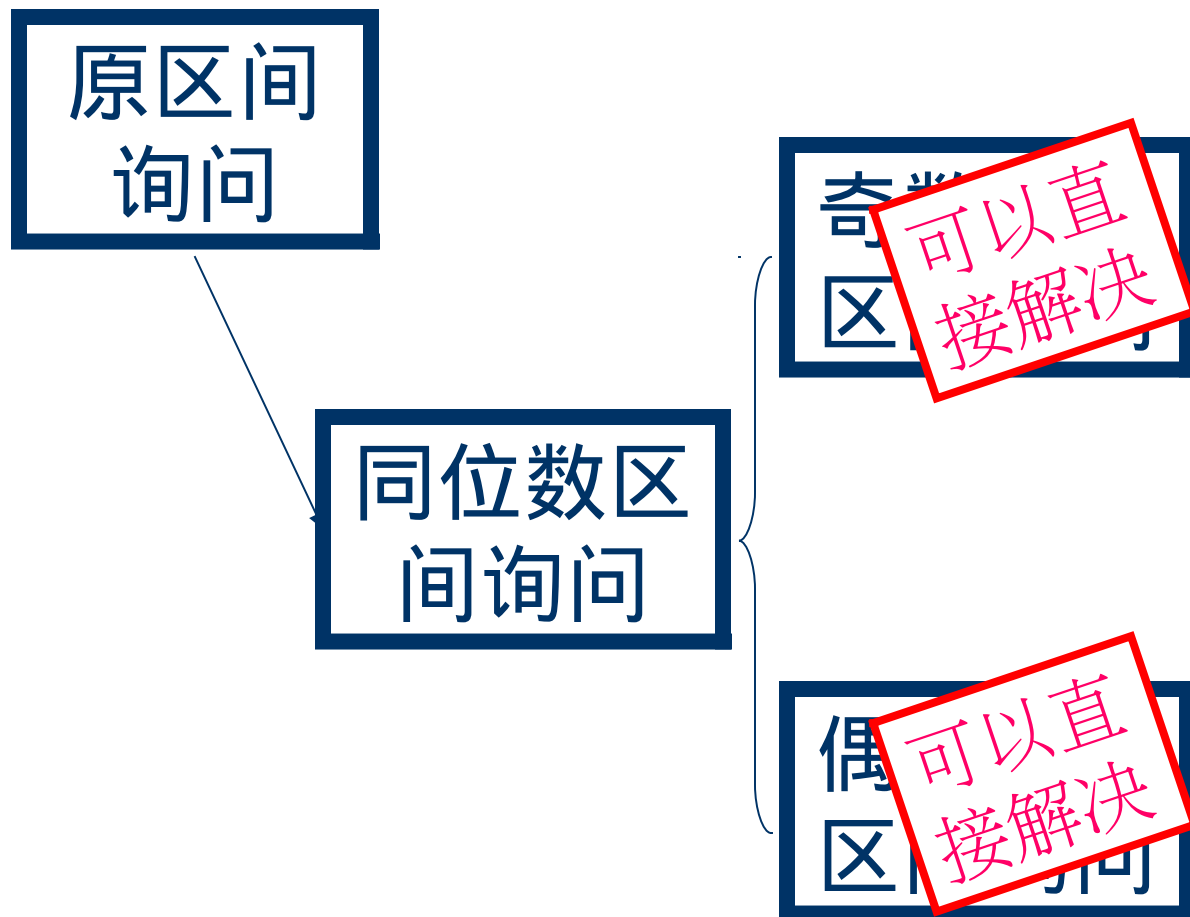
- $+1-0+0-0+0-3$

- $\dots\dots\dots$

- $+4-5+6-7+8-9$

可以看到，每一列的符号都是固定的，因此只需要对每一列分别进行求和即可

例题 1 总结



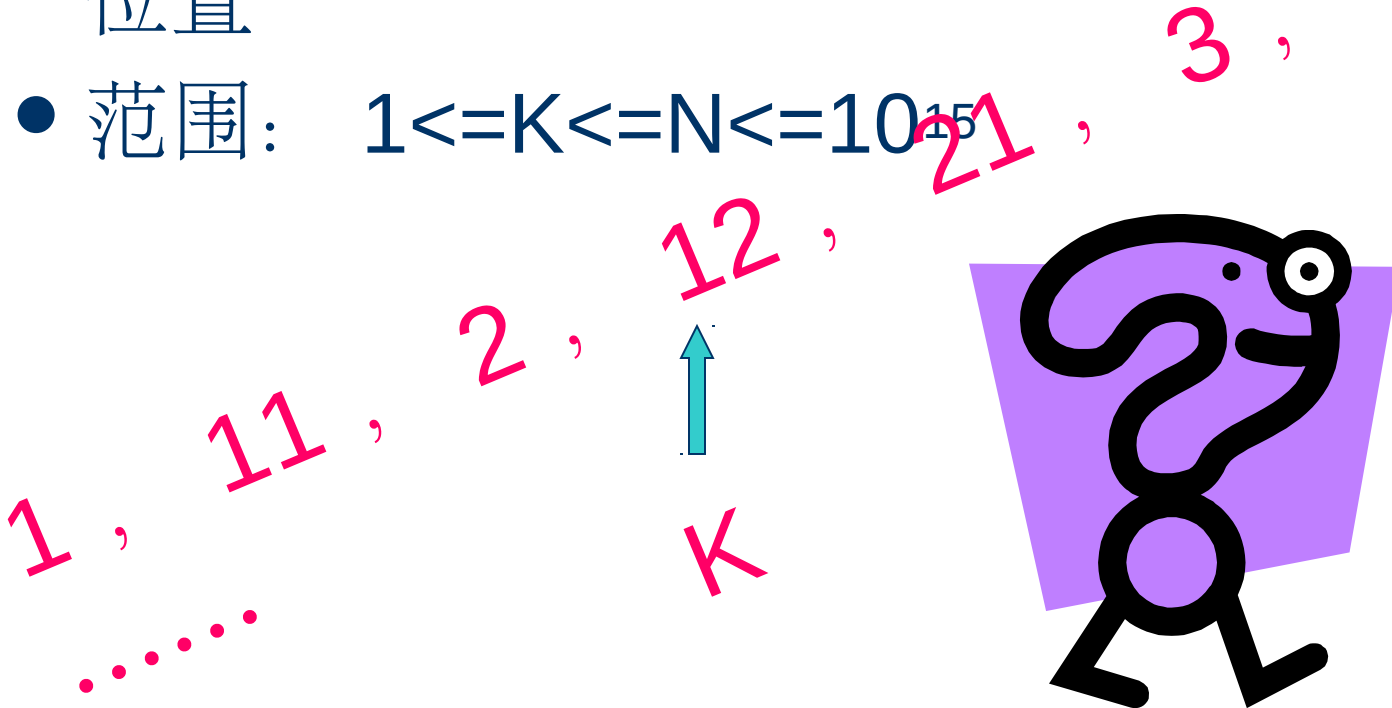
算法的本质是将复杂的问题逐步划分为简单问题的并，这正是解决数位计数问题的核心思想

例题 2 : Graduated Lexicographical Ordering (ZOJ 2599)

- 定义两个数的大小比较方法为首先比较各位数字之和，如果不相等则和大的数比较大，否则按字典序比较两个数的大小关系。
- 例如 120 小于 4，因为 120 的数字之和为 3，而 4 的数字之和为 4。555 小于 78，因为在字典序意义下 "555" < "78"。20 小于 200，因为在字典序意义下 "20" < "200"

例题 2 : Graduated Lexicographical Ordering (ZOJ 2599)

- 求 $1 \sim N$ 中第 K 大的数; K 在 $1 \sim N$ 中的位置
- 范围: $1 \leq K \leq N \leq 10^{15}$



例题 2：算法分析

- 原问题内有两问，事实上两问之间可以互相转化，因此首先考虑解决较为容易的一问。
- 对于原问题的两问，事实上似乎求 K 在 $1 \sim N$ 中的位置较为容易求出，因为它比较符合我们的解题思路。
- 我们可以将求 K 在 $1 \sim N$ 的位置换一种方式提出，即求 $[1, N]$ 中有多少个数比 K 小，这个问题可以通过区间划分的方法转化为更小的问题并加以解决。

例题 2：算法分析

- 尝试分解区间，我们发现，似乎怎样将区间拆分都不能将问题简化。
- 原因在于，对于比较两数的首要元素——数字和，在任何连续区间内，都没有很好的规律，可以直接利用。
- 那么，不妨转化思路，首先固定数字和，进而简化并解决问题。

例题 2：算法分析

- 首先固定数字和，问题转化为在区间 $[A,B]$ 内所有数字和为 S 的数当中，有多少个小于 K 。
- 显然，当 K 的数字和大于 S 时，答案等于 $[A,B]$ 区间内所有数字和为 S 的数的总数，当 K 的数字和小于 S 时，答案等于 0。
- 于是，我们只需解决两者相等时的情况。

例题 2：算法分析

- 新问题：当 K 的数字和为 S 时，在 $[A, B]$ 中所有数字和为 S 的数中，有多少个比 K 小。
- 这样，在数字和相同的情况下，只需考虑字典序，我们成功的简化了问题。

例题 2：算法分析

- 那么，下一步的区间划分主要考虑字典序的因素，因此按照首位的不同数字进行划分。
- 这里以 $[345, 45678]$ 为例 ($K=2457$)：
- $[345, 45678] = [345, 399] \cup [400, 499] \cup \dots$
 $\cup [900, 999] \cup [1000, 1999] \cup [2000, 2999] \cup \dots$
 $\cup [9000, 9999] \cup [10000, 19999] \cup [20000, 29999]$
 $\cup [30000, 39999] \cup [40000, 45678]$

例题 2：总结



可以直
接解决

数字和为 S2

可以直
接解决

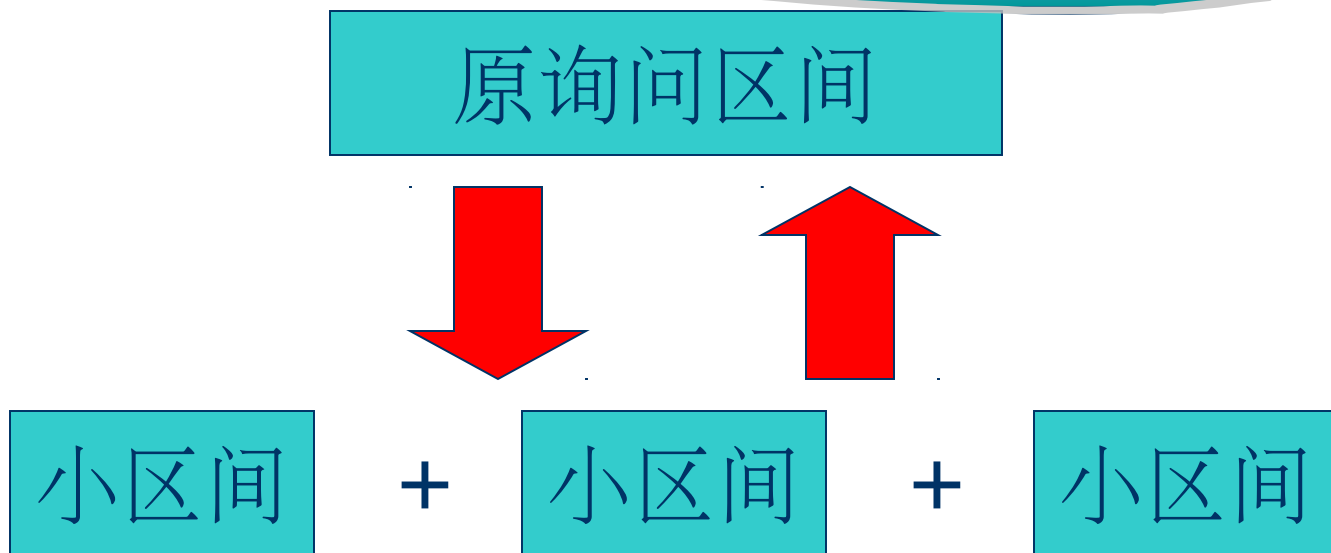
可以直
接解决

可以递归解决

可以直
接解决

总结

- 区间划分:



总结

- 适用范围：能够将大区间询问拆分为小区间询问结果的并。
- 若问题不满足上述条件，可以考虑对原问题加以转化使其能够满足以上条件。
- 区间划分的目的是简化问题，因此划分方式也要因题而异，具体问题具体分析。



Thank You!