

《网友小P》命题报告

安徽师范大学附属中学 曾致远

1 题意简述

求出有多少个字符串二元组 (S, T) ，满足如下条件：

- S 的长度为 N
- T 的长度为 M
- T 是 S 的一个子串
- S 的字符集大小为 A

2 数据范围

对于所有数据 $1 \leq M \leq 50, 1 \leq N \leq 200, M \leq N, 1 \leq A \leq 10^9$ 。

子任务编号	子任务分值	$M \leq$	$N \leq$	$A \leq$
1	3	50	200	1
2	5	8	8	2
3	7	20	20	2
4	5	1	200	10
5	10	1	200	10^9
6	10	12	12	10^9
7	25	50	50	10^9
8	35	50	200	10^9

3 算法介绍

3.1 子任务 1

$A = 1$ ：答案一定为 1。

3.2 子任务 2

$N \leq 8, A \leq 2$: $O(2^{N+M})$ 枚举所有字符串对后暴力判断即可。

3.3 子任务 3

$N \leq 20, A \leq 2$: $O(2^N)$ 枚举 S , 那么每个 S 对应的 T 的方案数就是 S 中本质不同的字符串数量。可以使用字符串哈希等方法计算。

3.4 子任务 4,5

$M = 1$: 考虑枚举 T 是哪个字符计算 S 的方案数。 A^N 是所有字符串的数量, $(A-1)^N$ 是不包括 T 当前枚举字符的方案数, 于是得到答案为 $A \times [A^N - (A-1)^N]$ 。

3.5 子任务 6

我们只需要枚举 S 的颜色集合, 而并不关心颜色具体是谁, 所以可以用最小表示法枚举 S 的颜色, 然后用一个排列数还原。之后用 Subtask 3 的暴力即可, 复杂度是 $O(\text{poly}(n)\text{Bell}(n))$ 。

3.6 子任务 7

这部分分数给实现不够优秀的标准算法。

3.7 标准算法

考虑枚举 T , 计算 S 的方案数。

设 $f(i)$ 表示 $|S| = i$, T 在后缀 $[i-M+1, i]$ 出现且这是第一次出现。转移枚举 T 并非第一次出现的方案数容斥, 有 $f(i) = A^{i-M} - \sum_{j=M}^{i-1} f(j) \times P(i, j)$, 其中 $P(i, j)$ 表示 T 同时在以 i 和 j 结尾的位置出现的方案数。

当两个对应的串不重叠时, 其中间的部分可以任意确定; 当两个对应的串重叠时只需要重叠的部分为 border 即可。

可以注意到, 两个串的 DP 转移完全相同当且仅当 border 集合完全相同, 长度为 n 的字符串合法 border 集合的数量为 $O(n^{\log n})$, 在本题中只有几千。搜出可行的 border 集合后将相同集合的 DP 一起计算即可。

4 数据生成方式

本题输入量小且为计数题，除了满足额外性质的子任务外，输入数据是否随机与做法无较大关联。所以数据一部分是随机生成，另一部分是恰好达到规定数据范围的上界。

5 命题总结

本题给了多档部分分，前5个子任务难度都不高，可以很轻松解决。

对于满分做法，从一个暴力的想法——枚举 T 入手、然后进行 DP，观察到关键的性质：DP 的转移只和 T 的 border 集合有关，很自然地想到暴力枚举 T 的 border 集合直接 DP，同时用容斥计算满足条件的 T 的方案数。这题一个很重要的关键点，就是当 $|T|$ 较小时，合法的 border 集合数量并不多，可以直接爆搜枚举，这是近几年在一些题目中出现的 trick，还是比较有意思的。本题作为一道偏难题。思路新颖，实现简单，解法比较自然，并留下了一定的拓展空间。你可以利用这一道题目，为你的集训队胡策之旅，画上一个完美的转折。