《多面体》解题报告

长沙市雅礼中学 段昱

November 28, 2018

1 题目来源

Balanced Strings, CS Academy

2 题目描述

对于一个仅由a，b，c组成的字符串S，我们称这个串是平衡的当且仅当，对于任意一个S的连续子串T，满足:

$$|f(T,x) - f(T,y)| \le K, \forall\{x,y\} \subset \{'a','b','c'\}$$

其中$f(T,x)$表示T中字符x的出现次数，K是给定的常数。

求有多少长度为n的平衡串，由于答案可能很大，输出它对P取模的结果。

3 子任务

对于所有数据，保证$1 \le n \le 10^9, 10^8 \le P \le 10^9, 1 \le K \le 8$.

- Subtask1(7%): $n \le 12, K \le 4$。
- Subtask2(15%): $n \le 10^3, K \le 4$。
- Subtask3(25%): $K = 2$。
- Subtask$i, i \in [4,8]$($5 \times 8\%$): $K = i - 1$。
- Subtask9(13%): 没有特殊的约束。

4 算法一

暴力枚举每个串，判断是否合法。

时间复杂度为$O(3^n \times n^2)$，期望得分7分。

5 算法二

设$cnt(i,x)$为字符x在$S[1...i]$（即S的第i个前缀）中的出现次数，那么如果S是平衡串，则对于任意一个子串$S[i...j], i \le j$，根据题意需要满足：

$$|(cnt(j,x) - cnt(i-1,x)) - (cnt(j,y) - cnt(i-x,y))| \le K, \forall\{x,y\} \subset \{'a','b','c'\}$$

不妨设$g(i,x,y) = cnt(i,x) - cnt(i,y)$，那么上面的条件就是

$$|g(j,x,y) - g(i-1,x,y)| \le K, \forall\{x,y\} \subset \{'a','b','c'\}$$

考虑递推，状态中记录当前的串长，以及之前$g(i,x,y)$的最大最小值，以及当前的$g(i,x,y)$的值分别是多少。转移时枚举下一个字符是什么即可。

复杂度为$O(nK^9)$或$O(nK^8)$，无用状态非常多，因此可以通过Subtask2。期望得分22分。

6 算法三

对算法二的递推状态稍作调整，记录之前$g(i,x,y)$的最大值与当前的$g(i,x,y)$的差，以及当前的$g(i,x,y)$与之前最小值的差，这样状态数和复杂度变为$O(nK^6)$。

由于这是一个线性递推，可以使用矩阵乘法来优化它。但计算发现矩阵大小达到了216，不过其中有不少状态是不可达的，且矩阵比较稀疏，稍作优化即能通过Subtask3。

这样复杂度为$O(K^{18}\log n)$，期望得分47分。

7 算法四

为了简化问题，先考虑只有两种字符时如何更高效地计算。此时字符串是平衡的，当且仅当:

$$\max\{g(i,a,b)|i \in [0,n]\} - \min\{g(i,a,b)|i \in [0,n]\} \le K$$

这样$g(i,a,b)$一定在一个长度不超过K的区间内，枚举这个区间，我们可以用矩阵乘法高效地计算有多少个串的$g(i,a,b)$在这个区间内。

但一个问题是怎么保证所有平衡串只被算一次。其实只需要先枚举所有长度为K的区间，用算出的答案之和减去所有长度为$K-1$的区间的答案之和就可以了。可以发现对于任意一个串，如果它的$g(i,a,b)$构成的区间长度为len，那么在第一步中会被算$K+1-len$次，在第二步中会被减去$K-len$次。

现在考虑原问题，对于一个平衡的串，我们需要保证$g(i,a,b),g(i,a,c),g(i,b,c)$分别满足在一个长度不超过$K$的区间内，类比上面的思路，我们可以分别枚举每个值的区间，然后进行容斥。设$F(l_1,l_2,l_3)$表示三个值构成的区间长度限制分别为l_1,l_2,l_3时算出来的答案之和，那么最终答案是：

$$F(K,K,K)-3F(K-1,K,K)+3F(K-1,K-1,K)-F(K-1,K-1,K-1)$$

注意这里用到了对称性，例如有$F(K-1,K,K)=F(K,K-1,K)$。

于是我们可以对每个F分开算。枚举三个区间后考虑矩阵乘法，一种比较暴力的做法是在状态中记录三个值，这样矩阵大小为$O(K^3)$。

由于我们需要做$O(K^3)$次矩阵乘法，复杂度为$O(K^{12}\log n)$，但矩阵比较稀疏，期望得分55-63分。

8 算法五

考虑优化算法四，由于有$g(i,a,c)-g(i,a,b)=g(i,b,c)$，状态中只需记录两个值，并根据第三个值的区间限制将一些非法状态去掉。矩阵大小降为$O(K^2)$。

这样做复杂度为$O(K^9\log n)$，期望得分87分。

9 算法六

上面的算法已经十分接近正解做法了，考虑继续优化。

假设我们在计算$F(K,K,K)$，枚举的三个区间分别是$[-x,K-x],[-y,K-y],[-z,K-z]$，考虑此时哪些状态$(u,v)$（表示$g(i,a,b)=u-x,g(i,a,c)=v-y$）合法：

$$-z\le(v-y)-(u-x)\le K-z$$

$$y-z-x\le v-u\le y-z-x+K$$

也就是说只要$y-z-x$相同，合法状态就是一样的，于是转移矩阵也一样。实现上只需枚举$y-z-x$，算出对应转移矩阵的n次幂即可。

最终复杂度为$O(K^7\log n)$，期望得分100分。