1 Decomposition

1.1 题目来源

Petrozavodsk Summer 2018. Day 8: Yuhao Du Contest 5 Problem D[1]

1.2 题目大意

对于字符串 T，记 $w(T)$ 表示 T 最短循环节的长度。

给定一个字符串 S，记一种划分为，将 S 分割为若干个非空字符串的拼接：$S = A_1 + A_2 + \cdots + A_m$。这种划分的权值为 $\prod_{i=1}^{m} w(A_i)$。

求 S 所有 $2^{|S|-1}$ 种划分的权值和，即：

$$\sum_{S=A_1+A_2+\cdots+A_m} \prod_{i=1}^{m} w(A_i)。$$

答案对 $10^9 + 7$ 取模。本题多测。

1.3 数据范围

对于所有数据，满足 $1 \le \sum |S| \le 10^6$。

1.4 解题过程

下述所有字符串下标从 1 开始。

考虑 DP，记 f_i 表示 $S_{[i,n]}$ 的所有划分的权值和。

枚举划分出第一个字符串的最短循环节 $S_{[i,i+x-1]}$。当然这里要求 $S_{[i,i+x-1]}$ 最短循环节长度为 x。

使用 SA 可以 $O(1)$ 找到最大的 k，使得 $S_{[i,i+kx-1]}$ 的最短循环节的长度为 x。则有转移：

$$f_i = \sum_{(i,x,k)} \sum_{j=1}^{k} f_{i+jx} \times x$$

该做法复杂度为 $O(n^3)$，难以通过。将上式进行一些数学上的转化，得到：

$$f_i = \sum_{j=1}^{n-i+1} f_{i+j} \times j + \sum_{(i,x,k)} F(i,x,k)$$

其中 $F(i,x,k) = \sum_{j=1}^{k} f_{i+jx} \times (i+x) - f_{i+jx} \times (i+jx)$。

这样转化的好处是：$k = 1$ 时 $F(i,x,k) = 0$，对转移没有贡献，可以直接忽略。

预处理出所有会对转移产生贡献的 (i,x,k)，值得注意的是，这样的三元组数量至多为 $O(n \log n)$ 个（后续将给出证明和求法）。

最后我们需要快速计算 $F(i,x,k)$，不妨考虑其与 $(i+x,x,k-1)$ 之间的关系。

[1] https://qoj.ac/problem/2206

F 的定义中两项形式非常相似，不妨拆开。

定义 $G(i,x,k)=\sum_{j=2}^{k} f_{i+jx}, H(i,x,k)=\sum_{j=2}^{k} f_{i+jx}\times(i+jx)$。则 $F(i,x,k)=G(i,x,k)-H(i,x,k)$。

根据两者的定义，有 $G(i,x,k)=G(i+x,x,k-1)+f_{i+2x}, H(i,x,k)=H(i+x,x,k-1)+f_{i+2x}\times(i+2x)$。

注意这里 j 的下标从 2 开始，这样 $k=1$ 时就有 $G(i,x,k)=H(i,x,k)=0$，易于用哈希表 $O(1)$ 维护。

下面来证明三元组数量的上界和给出找到所有三元组的方法：

首先证明，符合条件的三元组至多有 $O(n\log n)$ 个。事实上，对于固定的 i，这样的三元组至多只有 $O(\log n)$ 个，不妨令 $i=1$。

此时三元组合法的条件为：$S_{[1,x]}$ 最短循环节长度为 x，且 $S_{[1,x]}=S_{[x+1,2x]}$。

设所有合法的 x 构成的集合为 P，有结论：从 P 中任选三个元素 $a<b<c$，有 $\min(a+b,2b-a)\le c$。

证明考虑反证，选出 $a<b<c$ 使得 $\min(a+b,2b-a)>c$。由不等式，此时 $c-b<a$、$c-b+a<b$。接下来依次证明：

- 因为 $S_{[1,b]}=S_{[b+1,2b]}, S_{[1,c]}=S_{[c+1,2c]}$，所以 $c-b$ 是 $S_{[1,b]}$ 的周期。
- 因为 $S_{[1,a]}=S_{[a+1,2a]}$，所以 a 是 $S_{[1,\min(2a,b)]}$ 的周期。
- 因为 $c-b+a<\min(2a,b)$，根据周期引理（Periodicity Lemma），$\gcd(c-b,a)$ 是 $S_{[1,\min(2a,b)]}$ 的周期。
- 因为 $a<b$，所以 $\gcd(c-b,a)$ 也是 $S_{[1,a]}$ 的周期，又有 $c-b<a$，故 $S_{[1,a]}$ 的最短周期不是 a，与定义矛盾。

根据该结论，可以证明 $|P|=O(\log n)$：

- 将 P 中元素升序得到数组 $p_{1\sim m}$，记 $a_i=p_i-p_{i-1}$（$p_0=0$）。
- 设 $l_k=p_{k-1}=\sum_{i=1}^{k-1} a_i, r_k=p_{m-1}-p_k=\sum_{i=k+1}^{m-1} a_i$（$0<k<m$）。
- 找到 k 使得 $\max(l_k,r_k)\le\dfrac{p_m-a_m}{2}$，则 $a_m\ge\max(l_k,r_k)$。
- 若 $a_k\le\dfrac{p_m}{4}$ 或 $l_k\ge\dfrac{p_m}{4}$，则 $a_m\ge\dfrac{p_m}{4}$，将 a_m 删去后递归证明。
- 若 $a_k>\dfrac{p_m}{4}$ 且 $l_k<\dfrac{p_m}{4}$，则 $a_{k+i}\ge 2^{i-2}\times l_k$，则 $m-k=O(\log\dfrac{p_m}{p_{k-1}})$，将 $a_{k\sim m}$ 删去后递归证明。
- 有 $p_m\le n$，则 $|P|=O(\log n)$。

说明：此处仅需证明数量为 $O(\log n)$，证明过程中 $\min(a+b,2b-a)\le c$ 未必是最紧的限制，$\dfrac{p_m}{4}$ 也未必是最紧的下界。

接下来给出一个在 $O(n\log^2 n)$ 的时间内找到所有三元组的算法。

首先不考虑最短循环节长度这个限制，仅考虑表示出所有 $S_{[i,i+d-1]}=S_{[i+d,i+2d-1]}$ 的位置 (i,d)。

枚举 d，称下标为 kd 的位置为关键点（$1\le k\le\dfrac{n}{d}$）。此时 $S_{[i,i+d-1]}$ 一定经过某个关键点，而 $S_{[i+d,i+2d-1]}$ 经过下一个关键点。

枚举 $2 \le k \le \dfrac{n}{d}$，设关键点 $x=(k-1)d, y=kd$。使用 SA 求出 $S_{[1,x]}$ 和 $S_{[x+1,y]}$ 的最长公共后缀长度 p，$S_{[x,y-1]}$ 和 $S_{[y,n]}$ 的最长公共前缀长度 q。若 $d<p+q$，则 $x-p+1 \le i \le y+q-2d$ 的 i 满足 $S_{[i,i+d-1]} = S_{[i+d,i+2d-1]}$。以此法即可表示出所有 (i,d)，此技巧在《[NOI2016] 优秀的拆分》中亦有记载。

对 i 扫描线，用 set 维护出所有合法的 d。尝试据此升序求出所有 (i,x,k)。

若已经求出一组 (i,x,k)，则下一个合法 x' 满足 $\lfloor \dfrac{k}{2} \rfloor \times x < x'$ 且在 set 中出现。

必要性：$k<4$ 时即 $x<x'$；$k \ge 4$ 时 $\lfloor \dfrac{k}{2} \rfloor \times x$ 在 set 中出现，且不合法。

充分性：若存在 $x < x' < \lfloor \dfrac{k}{2} \rfloor \times x$ 合法，$S_{[i,i+2x-1]}$ 既有周期 x 也有周期 x'，则其有周期 $\gcd(x,x')$，即 $S_{[1,x']}$ 有长度小于 x' 的循环节。

此处的 set 可以换成常数更小的树状数组上二分。 仅此一处 $O(n\log^2 n)$，可以通过。

1.5 参考资料

无。