CC. PRDRAW *†

张晴川
qzha536@aucklanduni.ac.nz

December 13, 2020

大意

给一个字符串 S 和一个正整数 K。

对于每个 $1 \le L \le |S|$ 单独计算：随机选一个长度为 L 的子串 K 次，求 K 个子串两两不同的概率。

数据范围

- $1 \le |S| \le 200,000$
- $1 \le K \le 500$

题解

首先我们考虑对某个固定的 L 怎么计算。

假设长度为 L 的子串一共有 C 种，第 i 种的出现次数为 $\texttt{cnt}_i$，那么答案就是：

$$\sum_{i_1 \ne i_2 \ne \ldots \ne i_K} \frac{\texttt{cnt}_{i_1}}{N-L+1} \times \frac{\texttt{cnt}_{i_2}}{N-L+1} \times \cdots \times \frac{\texttt{cnt}_{i_K}}{N-L+1}$$

这个值等于：

$$K! \sum_{i_1 < i_2 < \ldots < i_K} \frac{\texttt{cnt}_{i_1}}{N-L+1} \times \frac{\texttt{cnt}_{i_2}}{N-L+1} \times \cdots \times \frac{\texttt{cnt}_{i_K}}{N-L+1}$$

由于 $K!$ 和 $N-L+1$ 的贡献都是固定的，我们只需考虑如何计算：

$$\sum_{i_1 < i_2 < \ldots < i_K} \texttt{cnt}_{i_1} \times \texttt{cnt}_{i_2} \times \cdots \times \texttt{cnt}_{i_K}$$

考虑多项式 $\prod_{i=1^C} (1 + \texttt{cnt}_i x)$，容易发现答案就是 K 次项的系数。

*https://www.codechef.com/problems/PRDRAW

†更多内容请访问：https://github.com/SamZhangQingChuan/Editorials

现在考虑如何求出每个 L 对应的 `cnt` 集合。

对 S 建后缀自动机。后缀自动机上每个节点对应的子串的长度是一个区间，所以对应的 `cnt` 影响的是一个长度区间。设区间为 $(a,b]$，考虑在 $a+1$ 处挂一个 $1+\texttt{cnt}x$，在 $b+1$ 处挂一个 $(1+\texttt{cnt}x)^{-1}$。

给定多项式 P，求 $P(1+cx)$ 是容易的，即 $a'_i = a_i + c \times a_{i-1}$。反过来求 $P(1+cx)^{-1}$ 的话，则是 $a'_i = a_i - c \times a'_{i-1}$。当 $c=1$ 的时候，实质上就是前缀和以及差分。注意到计算时不会依赖到比当前更高的项，所以截断到 x^K 即可，由此保证复杂度。

复杂度

- 时间：$O(NK)$
- 空间：$O(N+K)$

代码

https://gist.github.com/SamZhangQingChuan/23533064b1705149bb141c6ae1fd1023