

QOJ10545 String Rank 解题报告

题目大意

给定一个长度为 n ，仅由小写英文字母构成的字符串 s 。

定义 $Q_k(t)$ 为字符串 t 的所有长度 $\leq k$ 的子序列（包括空串）构成的集合（不可重集）。

求最小的 k ，使得 s 任意两个不同后缀的 Q_k 集合都不相同。

数据范围

$1 \leq n \leq 3 \times 10^6$ 。

解题过程

算法一

定义 $s[i:]$ 为字符串 s 中从第 i 个字符开始的后缀（ $s[n+1:]$ 为空串）。

限制等价于 $\forall i \in [1, n], Q_k(s[i:]) \neq Q_k(s[i+1:])$ 。

对每个 i ，计算是 $s[i:]$ 子序列但不是 $s[i+1:]$ 子序列的串的长度最小值 len_i ，那答案就是 len_i 的全局最大值。

考虑 s 是否有子序列 t 的判定方法：初始令 $cur = 0$ ，之后从前到后遍历 t_1, t_2, \dots, t_m ，在 s 的子序列自动机上跳 $cur = nex[cur][t_j]$ ，即：令 cur 变为 s 上 $> cur$ 的位置中第一个是 t_j 的位置。如果结束后 $cur \leq n$ ，则 s 中含有子序列 t ，否则不含有。

考虑使用该判定方法求出 len_i ：初始令 $cur1 = i - 1, cur2 = i$ ，然后依次遍历 t_1, t_2, \dots, t_m ，跳 $cur1, cur2$ 。结束后如果满足 $cur1 \leq n$ 且 $cur2 > n$ 则合法。

容易设计 dp ：令 $f_{i,j}$ 为初始 $cur1 = i, cur2 = j$ ，最终要跳到 $cur1 \leq n, cur2 > n$ 所对应 t 的长度最小值（无法跳到就是 INF）。转移枚举下一个选择的字符即可。

时间复杂度： $O(n^2|\Sigma|)$ 。

算法二

发现上述做法复杂度达到平方的原因是在两个串跳的过程中，需要维护两个指针 $cur1, cur2$ ，于是中间状态的数量达到了 $O(n^2)$ 。

如果对 $s[i:]$ 和 $s[i+1:]$ 从后往前的在反的子序列自动机上跑判定（ $pre[cur][t_j]$ 为 s 上 $< cur$ 的位置中最后一个是 t_j 的位置），就只需维护一个指针 cur ，初始 $= n + 1$ ，之后依次遍历 t_m, t_{m-1}, \dots, t_1 ，不断令 $cur = pre[cur][t_j]$ ，最后 $cur = i$ 则合法。

于是设计 dp ：令 f_i 表示从 $cur = n + 1$ 跳到 $cur = i$ ，所需 t 的长度最小值。转移枚举上一个选择的字符即可。

时间复杂度： $O(n|\Sigma|)$ 。

参考资料

QOJ10545 题面：<https://qoj.ac/problem/10545>

QOJ10545 题解：<https://qoj.ac/download.php?type=attachments&id=1989&r=1>