【CF1158F】Density of subarrays

【题目描述】

给定数列 a_n，定义 $f(a)$ 为最大的 p 使得任意 $x \in [c]^p$ 均为 a 的子序列，对于所有 $p \in [0,n]$，求 a 的 f 值为 p 的子序列个数，对 998244353 取模。

$n, c \le 3000$。

【思路】

Step #1：当 a 确定时，如何快速计算 $f(a)$？

Step #2：设 $f_{i,j}$ 表示以 i 为开头，f 值至少为 j 的子序列个数，最后差分即可，但是 f 如何转移呢？

Step #3：再预处理出一个 g 数组，其中 $g_{l,r}$ 表示开头为 l，结尾为 r 时，有多少子序列满足 $1 \sim c$ 全部出现并且 a_r 仅在 r 位置出现，即可快速处理 f，时间复杂度 $\mathcal{O}(\frac{n^3}{c})$。

Step #4：当 c 小的时候数据分治，将 $1 \sim c$ 中选过的数也压进状态即可。

【题解】

发现 $f(a)$ 等同于将 a 从前到后扫一遍，记一个集合 S，初始为空，每次扫到 $a_i = x$ 时，若 $x \notin S$，则将 x 加入 S，若加入后 S 的大小为 c，则将 S 清空，最终 $f(a)$ 等同于 S 被清空的次数，这个证明是显然的，由此还可以得到所有答案不为 0 的 p 都不超过 $\frac{n}{c}$。

于是就可以 DP，设 $f_{i,j}$ 表示以 i 为开头，f 值至少为 j 的子序列个数，再预处理出 $g_{l,r}$ 表示开头为 l，结尾为 r 时，有多少子序列满足 $1 \sim c$ 全部出现恰好一次并且 a_r 仅在 r 位置出现。

我们可以得到 $g_{l,r} = [a_l \ne a_r] 2^{\mathrm{cnt}_{a_l} - 1} \prod\limits_{i \ne a_l \cup i \ne a_r} (2^{\mathrm{cnt}_{a_l}} - 1)$，其中 cnt_x 表示 x 在 $[l, r]$ 中出现的次数。

而由此可以得出 f 的转移式：

$$f_{i,j} = \sum_{k=i+1} n g_{i,k} \sum_{l=k+1}^{n} f_{l,j-1}$$

用前缀和优化转移即可，时间复杂度 $\mathcal{O}(\frac{n^3}{c})$。

取一阈值 S，当 $c > S$ 的时候，直接用这个做法即可；当 $c \le S$ 的时候，把当前选了哪些数也压入状态即可，这样的时间复杂度是 $\mathcal{O}(n^2 \frac{2^c}{c})$。

取 $S = \log n$，得最终时间复杂度为 $\mathcal{O}(\frac{n^3}{\log n})$。