

《Far Away Song》解题报告

广州市第二中学 范璟阳

2026 年 10 月

目录

1	题目大意	2
2	数据范围	2
3	解题过程	2
3.1	Subtask 3	2
3.2	Subtask 7	3
3.3	Subtask 5	3
3.4	Subtask 6	3
3.5	Subtask 8 (正解)	3
4	总结	6
5	参考资料与致谢	6

1 题目大意

给定一个长度为 n 的字符串 S ，下标从 1 开始。记 $S(i, j) = S_i S_{i+1} \cdots S_j$ ，表示 S 从第 i 个字符到第 j 个字符组成的子串。

若非空字符串 B 同时是字符串 X 的前缀和后缀，且 $|B| < |X|$ ，则称 B 是 X 的一个 **Border**，其中 $|X|$ 表示字符串 X 的长度。

初始时，你位于子串 $S(1, n)$ 。每次跳跃按如下方式进行：

1. 在当前所在的子串中选择一个子串 X 。
2. 选择 X 的一个 Border B ，要求 $2|B| \geq |X|$ 。
3. 跳到 B 在 X 中作为前缀或后缀所对应的子串上，两者任选其一。

例如，当 S 为 aabaaba，你位于 $S(1, 7)$ 时，可以选择 $X = S(2, 7) = \text{abaaba}$ 。它的 Border aba 的长度恰好是 $|X|$ 的一半，因此你可以跳到 $S(2, 4)$ 或 $S(5, 7)$ 。

对于每个 $1 \leq i \leq n$ ，请分别求出从 $S(1, n)$ 出发、最终到达 $S(i, i)$ 时，最多可以进行多少次跳跃。若无法到达，则答案为 -1 。允许不进行任何跳跃。

2 数据范围

子任务	$n \leq$	$\sum n \leq$	特殊性质	分值
1	18	500	无	5
2	500			14
3	5×10^3	5×10^3		21
4	10^5	10^5		18
5	2×10^5	2×10^5	A	15
6			B	13
7			C	2
8			无	12

特殊性质 A：存在非空字符串 P ，使得 S 是无限循环串 P^∞ 的前缀，且 $|P| \leq 10$ 。

特殊性质 B： S 随机生成，即 S 的每个字符在大小 ≥ 2 的字符集中均匀随机。

特殊性质 C：对于所有 $1 \leq i < n$ ，均有 $S_i \neq S_{i+1}$ 。

对于所有数据，保证 $1 \leq n \leq 2 \times 10^5$ ，且单个测试点内所有测试数据的字符串长度之和 $\sum n \leq 2 \times 10^5$ 。字符串 S 仅包含小写英文字母。

3 解题过程

题目的跳跃操作中，选一个长度大于等于子串长度一半的 Border 保留，等价于选一个完整出现至少两次的周期，然后从子串的一端删掉它。

3.1 Subtask 3

考虑区间 dp，设 $f_{l,r}$ 表示从原串 $S(1, n)$ 跳到包含 $S(l, r)$ 的子串，最多能跳多少次。

初始化 $f_{1,n} = 0$ ，其余位置为 $-\infty$ 。有不计跳跃次数的转移 $f_{l+1,r} \leftarrow \max(f_{l+1,r}, f_{l,r})$ ， $f_{l,r-1} \leftarrow \max(f_{l,r-1}, f_{l,r})$ 。

结论 1：在跳跃操作中，只需要考虑选择最长的 Border。

证明: 同一个子串的所有前缀 Border 之间有包含关系, 后缀同理。换成同侧的最长 Border 后, 落点包含原落点, 原方案的后续操作仍然可以执行, 次数不会变少。

所以设 $S(l, r)$ 的最长 Border 长度为 b , 若 $2b \geq r - l + 1$, 则有转移

$$\begin{aligned} f_{l, l+b-1} &\leftarrow \max(f_{l, l+b-1}, f_{l, r} + 1), \\ f_{r-b+1, r} &\leftarrow \max(f_{r-b+1, r}, f_{l, r} + 1) \end{aligned}$$

在计算答案时不能直接拿 $f_{i, i}$ 作为答案, 因为题目的跳跃过程要求最终跳跃的终点是 $S(i, i)$, 而 dp 数组只限制了跳到包含 $S(i, i)$ 的子串上。

使用 $f_{i-1, i} + 1$ (若 $S_{i-1} = S_i$) 与 $f_{i, i+1} + 1$ (若 $S_{i+1} = S_i$) 中的最大值作为答案, 只考虑下标合法的项。若没有这样的项, 答案为 -1 。这是因为最后一步保留的 Border 长度为 1, 选取的子串只能长 2, 且两个字符相同。

固定左端点做 KMP, 就能得到所有右端点对应的最长 Border。按区间长度倒序转移, 复杂度为 $\mathcal{O}(n^2)$ 。

3.2 Subtask 7

有 $S_i \neq S_{i+1}$, 无法进行最后一步, 全部输出 -1 即可。

3.3 Subtask 5

设整个串 S 的最短周期为 p , 有 $p \leq 10$ 。若一次跳跃删除的周期比 p 大, 换成 p 后的落点一定包含原落点, 所以不妨只删除长度不超过 p 的周期。

我们可以对所有长度 ≤ 30 的区间跑区间 dp。

结论 2: 在当前区间长度 > 30 时, 不妨每次都从整个当前区间删掉一个长度为 p 的前缀或后缀。

证明: 假设选出的子串 X 的最短周期为 $q < p$ 。若 $|X| - q \geq p$, 则 $|X| \geq p + q$, 由周期引理, $\gcd(p, q)$ 也是 X 的周期。由于 q 最短, 得到 $q \mid p$ 。此时 X 包含一个完整的长度为 p 的块, 这个块又是长度为 q 的块的整次重复, 于是整个串也有周期 q , 矛盾。

所以一旦选取更小的周期, 落点长度一定 $< p$ 。当前区间长度 $> 30 \geq 3p$, 删去长度为 p 的前缀、后缀, 至少有一种仍然包含这个落点, 可以用它替换原跳跃。若仍然选择周期 p , 直接从整个当前区间删去同侧的 p 个字符, 也会包含原落点。因此可以一直删 p , 直到长度 ≤ 30 , 再进行其他跳跃。

只需要初始化长度在 $[31 - p, 30]$ 内的区间, 令 $dp(L, R) = \lfloor (L - 1)/p \rfloor + \lfloor (n - R)/p \rfloor$, 表示在两端尽量多删完整周期的次数。实际落点可以稍大于 $[L, R]$, 这里只要求包含它。其余初值为 0, 再跑短区间 dp。若 $n \leq 30$ 则直接用 Subtask 3 的做法。复杂度为 $\mathcal{O}(30n)$ 。

3.4 Subtask 6

S 随机生成, 于是我们猜想 S 中重复出现的串的长度不会太大, 对长度较小的区间跑区间 dp 即可。

在字符随机生成的模型下, 两个固定起点的长度为 t 的子串相同的概率至多为 2^{-t} , 最长重复子串长度以高概率为 $\mathcal{O}(\log n)$ 。

可以先用后缀数组求出最长重复子串长度 t (也可以直接令 t 为一个常数)。合法跳跃选出的子串长度一定不超过 $2t$, 所以只跑这些短区间的 dp, 初值全为 0 即可。复杂度以高概率为 $\mathcal{O}(n \log n)$ 。

3.5 Subtask 8 (正解)

结论 3: 存在一种最优跳跃方案, 使每一步删除的周期长度单调不升。

证明：上一步删掉周期 p 后，当前区间也有周期 p 。若下一步想删掉周期 $q > p$ ，则选出的子串长度至少为 $2q > 2p$ ，删除周期 p 同样合法，且同侧落点包含原落点。不断作这样的替换即可。

周期在被删除的时候一定是完整地重复出现了至少 2 次，这启发我们使用 Runs 来枚举周期。记一个 Runs 为 (l, r, p) ，表示 $S(l, r)$ 的最短周期为 p ，长度至少为 $2p$ ，且不能向左或向右延长并保持周期 p 。

根号分治，取阈值 B ，将跳跃的过程分成 3 步：

1. 进行若干步周期 $\geq B$ 的跳跃；
2. 选取一个周期 $p < B$ 的 Runs，从可达区间与它的交集中选出子串，不断从一端删掉 p 个字符，直到区间长度 $\leq 3B$ ；
3. 对所有长度 $\leq 3B$ 的区间跑区间 dp。

结论 4：存在一种最优方案，在进行完所有周期 $\geq B$ 的跳跃之后、当前区间长度降到 $\leq 3B$ 之前，只会用同一个小周期一直跳。

证明：设当前区间在最短周期为 $p < B$ 的 Runs 内。与结论 2 同理，如果下一步选取的子串 X 有更小的最短周期 $q < p$ ，则其落点长度 $|X| - q < p$ 。

当前区间长度 $> 3B > 3p$ ，删去长度为 p 的前缀或后缀，至少有一种能包含这个落点，所以可以将这一跳替换成删 p 。同周期的跳跃也可以改为从整个当前区间的一端删除 p 个字符。反复替换，直到进入长度 $\leq 3B$ 的区间即可。

3.5.1 第一步

设 $f_k(L)$ 表示在进行 k 次周期 $\geq B$ 的跳跃后，在所有左端点不大于 L 的可达区间中，最大的右端点。

初始化 $f_0(L) = n$ ，其余层没有可达区间的位置记为 0。注意 $f_k(L)$ 关于 L 单调不降，关于 k 单调不升。

转移考虑按 k 分层。对于一个周期 $p \geq B$ 的 Runs (l, r, p) ，枚举 $L \in [l, r - 2p + 1]$ ，令 $R = \min(f_k(L), r)$ 。若 $R - L + 1 \geq 2p$ ，则有转移 $f_{k+1}(L) \leftarrow \max(f_{k+1}(L), R - p)$ ， $f_{k+1}(L + p) \leftarrow \max(f_{k+1}(L + p), R)$ 。每层结束后对 L 取前缀最大值。

否则将这个 (L, r, p) 从后续的转移中删除，因为后面的 $f_k(L)$ 不会更大。

现在分析所有层的总复杂度。固定一个仍能参与转移的 (L, r, p) ，此时可达区间包含 $S(L, L + 2p - 1)$ 这个最短周期为 p 的平方。此前每次删除的周期都至少为 p ，否则这个平方也有更小的周期，与最短周期为 p 矛盾。所以这个候选至多参与 $\mathcal{O}(n/p)$ 层转移。

再由三平方串引理，同一起点的本原平方的根长每隔两项至少翻倍。因此固定 L 后，所有候选的 $\sum_{p \geq B} 1/p = \mathcal{O}(1/B)$ ，扫描总量为 $\mathcal{O}(n/B)$ 。第一步的总时间复杂度为 $\mathcal{O}(n^2/B)$ ，层数为 $\mathcal{O}(n/B)$ 。

3.5.2 第二步

由于之前的结论，我们只需要考虑长度 $> 3B$ 且周期 $p < B$ 的 Runs。

结论 5：这些 Runs 的总长度为 $\mathcal{O}(n)$ 。

证明：由周期引理，不同的两个 Runs 的重叠长度小于它们的周期之和，因而小于 $2B$ 。将每个 $[l, r]$ 缩短成 $[l, r - 2B]$ 后，所得区间两两不交，且保留的长度超过原长的三分之一，所以总长度为 $\mathcal{O}(n)$ 。

第二步每次只删 p 个字符，并在第一次长度 $\leq 3B$ 时交给第三步，因此只需要考虑长度在 $[3B - p + 1, 3B]$ 内的落点。

结论 6：这些 Runs 参与计算的左端点范围互不相交。

证明：一个 $\text{Runs}(l, r, p)$ 只需要处理 $l \leq L \leq r - 3B + p$ 的左端点，这个范围被 $[l, r - 2B]$ 包含，由结论 5 的证明可知它们互不相交。

现在固定左端点 L 与影响到 L 的 $\text{Runs}(l, r, p)$ ，只对 $R \in [L + 3B - p, L + 3B - 1]$ 且 $R \leq r$ 的区间初始化 $dp(L, R)$ 。

有一个暴力的想法是枚举这个 Runs 是从哪个区间开始往内跳的。设 $G(x, y)$ 表示只使用第一步的大周期跳跃，跳到包含 $[x, y]$ 的区间最多能跳几次，有 $G(x, y) = \max_{f_k(x) \geq y} k$ ，则第二步的贡献为

$$dp(L, R) \leftarrow \max \left(dp(L, R), \max_{l \leq x \leq L \leq R \leq y \leq r} \left\{ G(x, y) + \left\lfloor \frac{L - x}{p} \right\rfloor + \left\lfloor \frac{y - R}{p} \right\rfloor \right\} \right)$$

这里目标区间长度大于 $2B > p$ ，所以两端的完整周期可以逐个合法地删掉。

结论 7：只需要考虑 $x \in [l, l + p - 1]$ 或 $y \in [r - p + 1, r]$ 的区间 $[x, y]$ 的贡献即可。

证明：若 $G(x, y) = 0$ ，直接将 $[x, y]$ 扩大成整个 Runs 即可，贡献一定不会变小。

否则考虑跳到包含 $[x, y]$ 的区间之前的最后一步。设这一步删掉的周期长度为 q ，有 $q \geq B > p$ 。如果删的是前缀，那么上一个区间一定包含 $[x - p, y]$ ；如果删的是后缀，那么上一个区间一定包含 $[x, y + p]$ 。

只要 $x \geq l + p$ 且 $y \leq r - p$ ，这次扩大就仍然在 Runs 内。撤销最后一次大周期跳跃，再进行一次小周期跳跃，贡献不减。反复扩大，直到碰到上述两个边界之一即可。

先考虑 $x \in [l, l + p - 1]$ 的贡献。固定 L, y ， $G(x, y)$ 随 x 单调不降，而 $\lfloor (L - x)/p \rfloor$ 在这段区间内只有两种取值，每种取值只需要考虑最右边的 x 。

令 $x_0 = l + (L - l) \bmod p$ ，只需要考虑 $x = x_0$ 与 $x = l + p - 1$ ，后者要求 $l + p - 1 \leq L$ 。

固定其中一个 x ，枚举第一步跳跃次数 k ，最优的右端点为 $y_k = \min(r, f_k(x))$ ，且要求 $f_k(x) \geq R$ 。令 $t_k = r - y_k$ ， $A_k = k - \lfloor t_k/p \rfloor$ ， $c_k = t_k \bmod p$ ，则

$$k + \left\lfloor \frac{y_k - R}{p} \right\rfloor = \left\lfloor \frac{r - R}{p} \right\rfloor + A_k - \lfloor (r - R) \bmod p < c_k \rfloor$$

只需要维护最大的 A_k ，以及达到这个最大值时最小的 c_k 。其他 A_k 至少小 1，不会更优。倒序枚举 R ，由于 $f_k(x)$ 随 k 单调不升，用指针维护可行的 k 即可。

再考虑 $y \in [r - p + 1, r]$ 的贡献。设 $h_L(y) = \max_{l \leq x \leq L} \left\{ G(x, y) + \left\lfloor \frac{L - x}{p} \right\rfloor \right\}$ ，同样利用 $G(x, y)$ 随 x 单调不降，有 $h_L(y) = \max(G(L, y), h_{L-p}(y) + 1)$ ，其中第二项要求 $L - p \geq l$ 。

只需要保存 $y = r - p + 1, \dots, r$ 对应的 p 个值。对于每个 L ，用单调指针从 $f_k(L)$ 求出这些 $G(L, y)$ ，再通过上面的递推求出 $h_L(y)$ 。

固定 L, R 后，还需要最大化 $h_L(y) + \lfloor (y - R)/p \rfloor$ 。 $h_L(y)$ 随 y 单调不升，而 $\left\lfloor \frac{y - R}{p} \right\rfloor$ 在这段区间内也只有两种取值，所以每种取值只考虑最小的 y 。

只需要考虑 $y_0 = r - (r - R) \bmod p$ 与 $y_1 = r - p + 1$ ，将这两个位置的贡献用于更新 $dp(L, R)$ 即可。

记第一步的层数为 $m = \mathcal{O}(n/B)$ 。每个 L 的两部分转移都可以在 $\mathcal{O}(m + p)$ 内完成，而有效左端点范围互不相交，所以第二步总时间为 $\mathcal{O}(n^2/B + nB)$ 。

3.5.3 第三步

现在只剩下长度 $\leq 3B$ 的区间 dp ，状态含义与 Subtask 3 相同。初值全为 0，因为不跳跃时原串就包含任意区间。

还要补上第一步直接进入短区间的贡献。对于每个 L, k , 令 $R = \min(f_k(L), L + 3B - 1)$, 若 $R \geq L$, 则用 k 更新 $dp(L, R)$, 这一部分共需要 $\mathcal{O}(n^2/B)$ 时间。第二步的贡献也在处理对应的 L 时加入。

按左端点递增的顺序转移, 可以滚动数组优化, 空间为 $\mathcal{O}(B^2)$ 。

保存所有 $f_k(L)$ 需要 $\mathcal{O}(n^2/B)$ 空间, 取 $B = \Theta(\sqrt{n})$, 总时间、空间复杂度均为 $\mathcal{O}(n\sqrt{n})$ 。对 $f_k(L)$ 压位能让空间除一个 ω , 不过价值不大。

4 总结

本题题面简洁, 是一道较为困难的字符串题, 思维难度与代码难度均较高, 考察了选手对 Runs 性质的深刻理解。

本题设置了丰富的部分分, 其中 Subtask 3 与 Subtask 5 对正解有着强引导作用。

5 参考资料与致谢

本题是原创题, 没有参考资料。

感谢陈翔乐学长与 ChatGPT 对我的帮助。