《第二基地》解题报告

题目大意

给定整数 m，定义字符集 Σ 为前 m 个小写字母，对于两个字符集为 Σ 的串 A, B，定义 $f(A, B)$ 为如下问题的答案：存在一个有限大小的自动机 M，使得输入字符集为 Σ 的，任意长度的字符串 S，都可以比较 A 与 B 在 S 中的出现次数（返回 <，=，>）。如果存在，则 $f(A, B) = 1$，否则 $f(A, B) = 0$。给定 n 个串 $s_1 \sim s_n$，你要求出 $\sum\limits_{1 \le i < j \le n} f(s_i, s_j)$

数据范围

对于所有测试点，$2 \le n \le 10^6$，$N = \sum\limits_{i=1}^{n} |s_i| \le 10^6$，$2 \le m \le 26$。

子任务编号	$N \le$	特殊性质	分数
1	1000	$\|s_i\| \le 3$，$m \le 3$	10
2	5000	$m = 10$	10
3	10^6	$m = 10$	20
4	500	无	20
5	5000	无	10
6	10^6	无	30

解题过程

子任务 4

欲求 $f(A, B)$，考虑对 A, B 建 AC 自动机，一个点的点权为在这个点上 A 的出现次数减去 B 的出现次数，定义 $g(S)$ 为 S 中 A 的出现次数减去 B 的出现次数。

结论：$f(A, B) = 0$，当且仅当 AC 自动机上同时存在正环和负环，同时等价于对于任意整数 k（可以为负），存在一个串 S 使得 $g(S) = k$

充分性：假设能构造 l 个点的自动机 M，不妨假设 $|A| \ge |B|$，令 $1 \le p \le l + 1$，$q = l + |A| + |B| + 2$，如果同时存在正负环，可以构造两个串 S_1，S_2 满足 $g(S_1) = p, g(S_2) = -q$，令 $T = Concat(S_1, S_2)$，设 $h(p) = \min\{pos | g(T[1, pos]) = 0, |S_1| < pos \le |T|\} - |S_1|$，显然对于不同的 p，$h(p)$ 互不相同，而自动机只有 l 个节点，在加入完 S_1 后必有两个 p 位于同一状态，无法区分，得到的 $h(p)$ 相同，矛盾。因此同时存在正负环时不能构造。

必要性：如果 AC 自动机上没有负环，那么存在 V 使得 $g(S) \ge -V$，我们将 AC 自动机的每个节点 p 拓展为 $(g(S), p)$，当 $g(S) > V + |A| + |B|$ 时可以认为 A 次数总是比 B 多，因此自动机状态数有限。没有正环同理。■

有了这个结论，我们可以枚举两个串做一遍，时间复杂度 $O(n|S_i|^2)$，可以通过子任务 4，SPFA 可能可以通过子任务 5。

子任务 5,6

对只有两个串的 AC 自动机跑负环，很可能有特殊性质。通过推理和打表验证，可以发现，不存在串使得 B 比 A 多出现任意次，等价于满足下面两个条件之一。

条件一：A 在 B 中出现。

条件二：$m=2$，A 由 k 个 `b` 和 1 个 `a` 拼接，B 不全为 `b`，且结尾至少有 k 个 `b`。即 A 形如 `bbb...ba`，B 形如 `...bbb...b`。同理于 $A=$ `aaa...ab`，`baaa...a`，`abbb...b`。

证明：条件一显然，若不满足条件一，当 $m=2$，$A=$ `bbb...ba`，时，若 B 结尾没有 k 个 `b`，那么可以构造 $S=B$`a`B`a...`，将不会出现 A。若 B 结尾有 k 个 `b`，且 B 不全为 `b`，那么出现一个 B 后第一次出现 `a` 时将出现一次 A，因此 B 次数无法比 A 多。A 属于另外三种 pattern 同理。

接下来考虑 $m=2$，不满足条件一且 A 没有条件二中的四种 pattern。定义 $\neg c$ 表示对字符 c 取反，`a` 变为 `b`，`b` 变为 `a`。考虑对 A 做 KMP，假设匹配完 B 后 A 匹配到了 $pos<|A|$，接下来加入 $\neg A_{pos+1}$，显然 pos 不会变大，如果 pos 不变，说明 $A[1,pos]$ 每个字符都是 $\neg A_{pos+1}$，由于 A 没有 `bbb...ba` 这种 pattern，因此此时 $pos+1<|A|$。我们依次加入 A_{pos+1} 和 $\neg A_{pos+2}$，此时 pos 将会降至不超过 1。如果现在 pos 就是 1，由于 A 没有 `baaa...a` 这种 pattern，我们可以匹配到下一个和 A_1 相等的字符之前，然后再加入 $\neg A_1$，此时就可以将 pos 归 0 了，再加入一个 B 重复即可。

最后是 $m>2$ 的情况，此时条件一不变，如果不满足条件一，在匹配完 B 后 A 匹配到了 pos，如果 $pos>0$，可以加入不同于 A_{pos}，A_{pos+1} 的第三种字符使得 pos 严格变小。因此 $m>2$ 时原命题就等价于满足条件一。■

对于统计答案，$f(A,B)$ 同时没有正负环的情况只有两种，第一种是 $A=B$，第二种是 $m=2$，$A=$ `abbb...b`，$B=$ `bbb...ba`（翻转同理），不难由前面的结论推出。注意两种条件都要求 $|A|\le|B|$，因此都满足只能是 $|A|=|B|$，同时满足条件一或同时满足条件二。剩下情况至少有正环或负环，那么直接算没有正环和没有负环的总个数即可。

对于条件一，使用你喜欢的字符串算法即可。例如 AC 自动机，将每个串插入到 AC 自动机上，答案就是每个串前缀对应节点在 fail 树上的虚树的权值和。

对于条件二但不满足条件一的情况，将前面四种 pattern 的个数统计一下，每个串要求这个 pattern 的长度在中间最长连续 `a`与最后一段 `a` 长度之间，前缀和一下即可。总时间复杂度 $O(n\log n+nm)$

参考资料&后记

感谢王泽远学长提供本题的 $n=2$，$|S_i|\le 5000$ 版本，即子任务 4 的做法。他认为 $f(A,B)$ 应该有更多性质。笔者继续推理和验证后得到子任务 6 的做法。直接猜条件一可以通过子任务 $2,3$。考虑到本题的精髓在 $m=2$，前面 3 个子任务只给了较少的分数。