

Chef and Rectangle Genome解题报告

长郡中学 任瀚林

1s,256MB

§1 题面

§1.1 问题描述

(注：这是codechef上的官方翻译)

给定一个 $N \times M$ 的空矩阵，大厨希望填上G,C,A,T以构成一个尽可能稳定的基因矩阵。

基因矩阵的稳定性与其互不相同的子矩阵的数目相关——越接近 K 越好。

请帮助大厨构建尽可能稳定的基因矩阵。注意，你必须在所有的空位置填上字符，两个子矩阵被认为是不同的，如果这两个子矩阵拥有不同的大小或者至少有一个相同的位置上的字符不相同。

§1.2 输入格式

输入数据的第一行包含一个整数 T ——测试数据的组数。

对于每组测试数据，仅包含一行三个整数 N, M, K 。

§1.3 输出格式

对于每组测试数据，输出一个 $N \times M$ 的仅包含G,C,A,T的矩阵。

§1.4 样例输入

```
2
2 3 5
1 1 1
```

§1.5 样例输出

```
AAA
AAA
T
```

§1.6 评分方式

对于一组测试数据，设你返回的答案有 W 个互不相同的子矩阵，这组数据的“得分”将是 $\frac{|W-K|}{K}$ 。总的“得分”是所有数据“得分”的和。

对于一个测试点，假设我们给出的解的“得分”为 std ，你给出的解的“得分”为 you ，那么你的分数为

$$5 \times \left(\frac{std}{you} \right)^{2.333}$$

std 的值会比较弱。

§1.7 数据规模和约定

测试点比例	$T =$	$N \leq$	$M \leq$
5%	10	3	3
10%	20	5	5
10%	40	7	7
15%	100	9	9
20%	100	11	11
20%	100	13	13
20%	100	15	15

数据生成方式为: N, M 是人为构造的, K 是在 $[1, N^2 M^2]$ 内均匀随机的。(具体见原题)

§2 解题报告

考虑给出一个矩阵 A , 如何求出其不相同的子矩阵个数 $\text{ans}(A)$ 。直接用矩阵哈希可以做到 $O(N^2 M^2)$ 的复杂度。

我们可以使用一种贪心方法生成一个 $\text{ans}(A)$ 很接近 K 的矩阵。一开始 A 是全 A 矩阵, 逐位随机产生 $A_{i,j}$ 的值, 如果 $\text{ans}(A) > K$ 则把 $A_{i,j}$ 赋为 0。这个做法复杂度是 $O(N^3 M^3)$ 的, 无法承受。但是可以用二分思想, 每次把剩下的 0 中的一半生成, 如果 $\text{ans}(A)$ 超过了 K , 那么就把这一半赋为 0。

但是这样不一定是最好的。我们可以考虑另外一种方法: 每次确定一个元素 $A_{i,j}$, 只考虑所有元素都已经被确定的矩阵中, 不同的矩阵个数。确定一个元素后可以 $O(NM)$ 地算出贡献。枚举当前这个元素是 ATCG 的哪一位, 估算出应该有的不同矩阵个数 $k_{i,j}$, 取最接近 $k_{i,j}$ 的作为这一位。

一个较好的 $k_{i,j}$ 为 (前提是: 如果 (i, j) 已经被确定了, (k, m) 没有被确定, 那么 $\neg(k \leq i \wedge m \leq j)$)

$$k_{i,j} = \frac{K \sum_{i \leq n, j \leq m} ij}{\sum_{(i,j) \text{ 已被确定}} ij}$$

当然还有其他的求出初始解 A 的方法。不同的方法精确度不同, 效率也不同。

由于生成的初始解一般不是很优, 我们还要考虑随机调整我们的解。每次随机改变 A 中一个元素, 如果能够导致更优解产生, 那么直接接受该解; 否则不接受 (即爬山) 或者以一定概率接受 (退火?)。调整的次数也有讲究, 一般来说与 $TN^2 M^2$ 成反比, 有时候可以人为改变这个参数 (实际运行时间较小的时候, 或者 K 比较小的时候, 调整次数变大)。

这样就可以取得不错的分数了。

§3 数据生成

N, M 是人工确定的 (也有随机产生的), K 随机地在 $[1, N^2 M^2]$ 中生成。

§4 参考资料

[1] 官方题解: <https://discuss.codechef.com/questions/47242/gerald09-editorial>