

《友好的生物》解题报告

【摘要】

基本思路	枚举+扫描
时间复杂度	$O(N \log_2 N + 2^K \cdot NK)$
空间复杂度	$O(NK)$

【题意简述】

共有 N ($2 \leq N \leq 100,000$) 种生物，对于任意的生物 t ($1 \leq t \leq N$)，都有 K ($1 \leq K \leq 5$) 种属性： $P_{t,1}$ 、 $P_{t,2}$ 、 \dots 、 $P_{t,K}$ 。此外，根据给定的非负常数 C_1 、 C_2 、 \dots 、 C_K ，利用以下公式，可以计算出任意两种生物 A 和 B 之间的友好程度：

$$Friendline\ ss = \left(\sum_{i=1}^{K-1} C_i \times |P_{A,i} - P_{B,i}| \right) - C_K \times |P_{A,K} - P_{B,K}|$$

现在给出各种生物的属性值，要求找出友好程度值最大的一对生物。

【算法分析】

根据题中给出的公式，首先可以得出一个对问题的简化：开始时，将每种生物的第 i 种属性都乘以 C_i 。这样，在之后的计算中，就不需要再考虑 C_i 数列，

因而得到了新的友好程度计算公式： $Friendline\ ss = \left(\sum_{i=1}^{K-1} |P_{A,i} - P_{B,i}| \right) - |P_{A,K} - P_{B,K}|$ 。

由于公式中，最后一种属性差值的系数为-1，与前面 $K-1$ 不同。为了方便分析，我们首先只考虑前 $K-1$ 种属性差值，也就是说，两种生物的友好程度恰好为它们各项属性差值的和。我们可以对新的公式进行如下变形：

$$F' = \sum_{i=1}^{K-1} |P_{A,i} - P_{B,i}| = \sum_{i=1}^{K-1} S_i \times (P_{A,i} - P_{B,i}) = \sum_{i=1}^{K-1} S_i \cdot P_{A,i} - \sum_{i=1}^{K-1} S_i \cdot P_{B,i}$$

其中， S_i 是根据 $P_{A,i}$ 与 $P_{B,i}$ 的大小关系给出的系数：当 $P_{A,i} \geq P_{B,i}$ 时， $S_i=1$ ；反之，当 $P_{A,i} < P_{B,i}$ 时， $S_i=-1$ 。因此，对于确定的序列 S ，任意的生物 t 将会有一个特征值 $E_t = \sum_{i=1}^{K-1} S_i \cdot P_{t,i}$ 。

那么，对于任意两种生物 A 和 B ，可以首先根据两种生物各项属性的大小关系，得出合适的序列 S ，进而得出 E_A 和 E_B ，进而可以确定它们之间的友好程度，将恰好等于 $E_A - E_B$ 。

很显然，问题的难点就在于，可能的 S 序列共有 2^{K-1} 种，它将随着生物种类的变化而变化，无法预先确定。但通过进一步分析可以发现，实际上，对于生物 A 和 B ，根据合适的序列 S 得出的友好程度 $E_A - E_B$ ，将比根据其他各种序列 S' 得出的友好程度都要大。因为根据合适的序列 S ，最后的友好程度将等于一系列非负整数的和；而如果改变序列 S 中若干项的符号，也就相当于将原来求和式中，对应的非负整数改为非正数，相应的和也就不会增加。

也就是说，对于两种生物，它们之间真正的友好程度，将等于各种符号序列所推出的友好程度中最大的一个。因此，我们要求所有生物中，友好程度最大的一对，也就是根据每种符号序列分别推出的生物特征值 E_i 中差距最大的一对。

这样，我们得出了在只考虑前 $K-1$ 种属性的前提下，求友好程度最大的一对生物的方法：枚举 2^{K-1} 种符号序列，每次根据确定的符号序列 S ，计算出所有生物的特征值 E_i ，并找出其中差距最大的一对（特征值最大的和最小的两种生物）。最后的答案也就是那 2^{K-1} 对生物中，特征值差最大的一对。

解决了上面的问题，我们就可以将第 K 种属性也考虑进来。由于最后一项属性差值的系数是 -1 ，如果仍然沿用前面的方法，枚举出 2^K 个符号序列，将不能保证生物之间真正的友好程度是根据各种 S 序列得出的友好程度中最大的一个。

下面，我们的目标就是要消除对第 K 种属性值符号的不确定性。我们注意到，公式中最后一项属性差值的系数为负值，也就相当于每次都用较小的一个属性值减去较大的。而对于两种生物，一旦确定将第 K 项属性差值的相反数（用较小的属性值减去较大的所得结果）计入最后的友好程度，那么前面 $K-1$ 项属性差值的和，仍可以根据前面的方法，通过枚举符号序列求得。

综合上面的分析，我们就得出了下面的方法：枚举前 $K-1$ 项属性的符号序列，同时不妨假设 $S_K=1$ ；根据确定的符号序列 S 求出每种生物的特征值 E 。那么生

物 t 与其他生物的最大友好程度就等于：
$$F_{\max} = \text{Max} \begin{cases} E_i - E_t (P_{i,K} \leq P_{t,K}) \\ E_t - E_i (P_{i,K} > P_{t,K}) \end{cases}$$
，由

此就可以得出所有生物种友好程度最大的一对生物。

到此，我们已经基本得出了解决本题的算法。其中比较复杂的部分就是在确定符号序列之后，根据上面的式子，对于每种生物，快速的找出与其他生物的最大友好程度。我们可以利用线段树、AVL 等一些支持高效率的插入和求最值操作的数据结构解决这个问题——对于每种生物 i ，首先在现有集合中查找与它友好程度最大的生物，然后将生物 i 插入集合，以便后面其他生物的查找。这样的方法较为复杂，仍然需要进一步的优化。

我们注意到，求最大特征值差的式子中，每种生物都要和其他所有生物比较。而实际上，在求最大友好程度的过程中，根据大小关系的“对称性”，完全可以不考虑比当前生物的第 K 种属性值小的生物，每次只用当前的生物的特征值，减去那些比自己第 K 种属性值大的生物的特征值，从这些结果中找到最大的一个。

【算法描述】

根据上面的分析，我们得出了下面的算法：

首先，将所有的生物，按照第 K 种属性值从大到小排列；

然后，枚举出各种符号序列（共 2^{K-1} 种），对于每一种符号序列 S ，求出每种生物的特征值 E ，接着依次扫描每种生物，扫描中记下前 $i-1$ 种生物最小的特征值 Min ，并且用 $E_i - Min$ 与当前最大的友好程度进行比较，并且用 E_i 改进 Min ，便于以后的扫描中进行比较。

【时空复杂度】

空间方面，该算法只需要记下每种生物的各项属性值，空间复杂度为 $O(NK)$ 。

时间复杂度方面，第一步排序，由于数据给出的属性值是 $[-10000, 10000]$ 的整数，因此可以用一些线性的排序算法，复杂度为 $O(N)$ ；更一般的情况下可以直接使用快速排序等，复杂度为 $O(N \log_2 N)$ 。第二步，一共存在着 2^{K-1} 种符号序列，每次扫描的复杂度均为 $O(NK)$ ，因此总的复杂度为 $O(2^K NK)$ 。因此整个算法的时间复杂度就是 $O(N \log_2 N + 2^K NK)$ 。

【解题小结】

回顾解决本题的过程，原来对于两种确定的生物，只可能得出一个友好程度，而我们却根据不同的符号序列计算出若干个友好程度值，使得原来具有差异的计算得到了统一，方便了计算。同时，由于这样的**统一**是建立在不影响最后结果的基础上的，使得整个问题得到了圆满的解决。

此外，在分析问题时，我们先从简单的情况入手，进而分析更加复杂的情况，这样的思路在今后解题过程中也有一定的借鉴意义。

【附录】

一、本题原题：[Species.doc](#)

二、源程序：[Species.pas](#)