

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会  
电话：0471-4969085

邮编：010021  
参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)  
Email: 2015@tzmcm.cn

---

## 第八届“认证杯”数学中国

### 数学建模网络挑战赛

### 承 诺 书

我们仔细阅读了第八届“认证杯”数学中国数学建模网络挑战赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们接受相应处理结果。

我们允许数学中国网站([www.madio.net](http://www.madio.net))公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

**我们的参赛队号为：**

**参赛队员（签名）：**

队员 1：何锦彪

队员 2：杨恪鸣

队员 3：封 烨

**参赛队教练员（签名）：** 方红

**参赛队伍组别（例如本科组）：** 本科组

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

---

## 第八届“认证杯”数学中国

### 数学建模网络挑战赛

### 编号专用页

参赛队伍的参赛队号：（请各个参赛队提前填写好）：

1134

竞赛统一编号（由竞赛组委会送至评委团前编号）：

---

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会  
电话：0471-4969085

邮编：010021  
参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)  
Email: 2015@tzmcm.cn

## 2015 年第八届“认证杯”数学中国 数学建模网络挑战赛第一阶段论文

题 目 荒漠区生态退化评估及治理研究

关 键 词 *SNN*、投影寻踪法、*RAGA*、*K-Means*、元胞自动机

### 摘 要：

目前，土地的荒漠化已经成为一个世界热点话题。土地的荒漠化会导致土地生产力的下降、土地资源的丧失以及对环境的破坏，对我们人类的发展与生存产生极大的影响。因此，研究对在荒漠地域生态退化程度的判断标准以及在此标准下在各阶段的补救措施具有重要的现实意义。

对于问题一，首先，根据分析可知，生态退化的程度可以在调查的动植物的种群状态上反映出来，故我们将植物因子与啮齿动物优势种百夹捕获率作为生态退化程度的评价指标，将附件中的重复项作为样本。同时我们发现数据中存在着部分的缺失，由于数据的时序关系较弱，插值等数据补充方法难以应用，故将存在缺失数据的样本删除。接着，为了减少数据偶然性的影响，我们利用基于共享型最近邻居相似度算法（*SNN*）进行了异常值的筛检，共去除了9个异常样本（见表 2）。然后，对数据进行归一化后，我们建立了投影寻踪模型，利用基于实数编码的加速遗传算法求解得到最佳投影方向为  $a^*=(0.0761, 0.1167, 0.1735, 0.1113, 0.2068, 0.3520, 0.2208, 0.1362, 0.5483, 0.5081, 0.3811)$ 。利用最佳投影方向计算各样本的投影值（见附录3），由于投影值能很好地表现高维数据的分类结构特征，故将其作为评价价值进行评价模型的构建。最后，为获得生态退化、半退化与未退化状态的区分标准，我们采用了*K-Means*聚类算法将样本的评价价值聚为3类，分析评价价值的属性，以生态状态来解释各分类类别，并以类间的中间距离作为不同状态的分界线得到区分标准。据此讨论不同人类活动造成的荒漠地区生态退化情况（见表 8），发现开垦会造成土地极大的破坏，在开垦的影响下，88.37%的地域处于生态退化状态，过牧对土地的负面影响次之，轮牧对土地的负面影响较于前两者而言相对较弱一些。

对于问题二，为了直观地反映过牧、轮牧、开垦三种荒漠地区的半退化和退化的动态变化情况，我们采用元胞自动机（CA）进行仿真模拟。在不改变当前人为干扰和人工植被的情况下，三种荒漠区呈现出不同的生态退化情况，轮牧区较轻，过牧区略重，开垦区非常严重。然后，分别考虑减少人为干扰和增加人工植被两种方法进行治理，结果发现轮牧区和过牧区在减少35%人为干扰或增加一倍人工植被的情况下呈现出良好的恢复趋势，而开垦区减少65%人为干扰也难有起色，只有增加3倍人工植被才能慢慢恢复。说明对于过牧区和轮牧区能有效治理，而开垦区的治理花费太大，需要寻找新的方法。

最后我们针对模型做了敏感性分析，并进行了中肯的评价和适当的推广。

参赛队号：1134

所选题目：C 题

参赛密码 \_\_\_\_\_  
(由组委会填写)

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会  
电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)  
Email: 2015@tzmcm.cn

## 英文摘要（选填）

（此摘要非论文必须部分，选填可加分，加分不超过论文总分的 5%）

## Abstract

In the present era, the desertification has become a hotly-discussed topic. The desertification may bring the decreased production, the lost of land resource and great damage to environment. The human life and development are greatly influenced by desertification. Therefore, it's of great significance to research on ecological degradation criterion and make some remedial measure.

Aiming at question one, firstly, according to the analysis, the degradation extent can be reflected on the investigated animal or vegetable population, the plant factor in the appendix and the capture rate of dominant rodent species are considered as the degradation index and duplicate item in each row was regarded as a sample. At the meanwhile, because of the number partial deficiency, the weak sequential relationship and hardness of applying supplementary data, deleting the deficient data sample. Then, for decreasing the contingency influence, the SNN based on shared "nearest neighbor" was employed to test outlier and totally eliminated 9 outliers (seen from chart). Then, after normalization, the Projection pursuit model and Accelerating Genetic Algorithm based on real-code were employed to get the optimal projection direction, following as  $a^*=(0.0761, 0.1167, 0.1735, 0.1113, 0.2068, 0.3520, 0.2208, 0.1362, 0.5483, 0.5081, 0.3811)$ . Utilize the optimal projection direction to calculate the projection value and regard it as evaluation value to construct the evaluated model result from it can perfectly present the high dimensional data features. Finally, for obtaining the degradation, semi-degradation and non-degradation distinguish criterion, the K-Means clustering algorithm was employed to congratulate the sample into 3 categories, analyzing and evaluating properties of value, explaining the different classes with ecological condition and getting distinguished criterion of ecology with the semi-distance seen as boundary. Discuss the degradation caused by human activity.(seen from chart) and discover that reclamation causing the great damage. Under the reclamation influencing, the 88.37% ground are declining. Overgrazing has a slight negative effects and the influence of rotational grazing is much slighter than the former ways.

According to question two, for intuitional reflecting the degradation and semi-degradation dynamic change of three desertification. The Cell Automation was employed to simulate. Under the condition that not changing the interference and artificial vegetation, three areas present different degradation, reclamation area is most serious and rotational grazing is lighter. Next, apply the decreased the interference and increased artificial vegetation to govern, the result shows that the rotational grazing and overgrazing areas improve with decreasing 35% interference and increasing one times artificial vegetation, moreover, reclamation area cannot optimize decreasing 65% interference and only by increasing 3 times artificial vegetation can recover. It indicates that it's expensive to govern reclamation and it's reasonable to restore the other two areas and it also shows finding another new pattern.

Finally, the sensitive analysis, pertinent evaluation and pertinent evaluation was conducted towards the model.

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

## 一、问题的提出

### 1.1. 研究背景

法国学者奥布雷维莱 (A.Aubreville) 提出荒漠化一词后，由于当时认识水平较低且学科发展较缓慢，对于荒漠化问题的研究尚未引起足够的重视。1977 年内罗毕联合国荒漠化大会后，荒漠化问题便成为一个世界热点话题，也是全人类所聚焦的关键问题。与此同时，荒漠化的评价问题也被提出来，世界上许多国家的学者都致力于这方面的研究。然而荒漠化问题之所以受到国际社会的广泛关注是因为它是直接威胁人类生活赖以生存的自然环境的一种灾害。这种灾害的直接导致土地生产力的下降、土地资源的丧失以及环境的极大破坏。研究荒漠化根本目的在于发现土地退化的趋势、如何有效地防御、减少和控制这种灾害的发生与蔓延，恢复其生态平衡。在联合国荒漠化的定义框架内，建立和完善荒漠化的理论体系，特别是荒漠化评价的指标系数，是荒漠化理论的核心和关键，同时也是荒漠化问题研究的热点<sup>[1]</sup>。

目前世界上许多国家都在积极探索，我国作为世界上受荒漠化危害最严重的国家之一，也急需建立科学、合理和生产实用性强的荒漠化评价指标体系，既能用于指导我国荒漠化防治的实践，同时又能与国家防治荒漠化行动接轨。荒漠化评价指标体系的研究经历了约 30 年的时间，但迄今为止还没有一个公认的、被广泛接受和推广性强的荒漠化评价指标体系<sup>[2]</sup>。分级标准科学根据不足，指标的临界值或量化值大多是人为确定，带有相当的主观性，难以客观反映和准确评价土地荒漠化实际状况。当前多数评价的目的性不明确，指标较繁杂，指标间的交叉性大，内容重复。从人类活动的角度出发，人类活动是一种生产活动，但对于自然生态系统来说，人类的所作所为均是一种干扰。人为活动对荒漠生态系统动植物的影响主要指受人为因素控制的土地开垦、放牧等活动对荒漠植物、动物数量及分布格局的影响。

本文在分析荒漠啮齿动物优势种群在不同干扰条件下数量动态变化特征<sup>[3]</sup>，从而进一步从干扰的理论角度来分析荒漠化的内在原因和规律。我们根据所给数据建立数学模型评估人类活动造成的荒漠地区退化程度。此外分析荒漠地区处于不同阶段时如何通过减少人为干扰或其他方法来促使该地区恢复正常。给出可行方案以及造成原因。

### 1.2. 文献综述

通过查阅资料，我们发现目前已给出荒漠化各种指标的评价方法，例如荒漠化遥感评价、荒漠化危险性评价及荒漠化发展速率评价等。大量的文献都是关注于荒漠化指标的评价，而且都是以宏观的定性指标。由于对荒漠化概念理解的差别，指标选取各不相同，可能性小，难以在地理分以复杂的范围应用。除此之外，评价的层次性不强，多数指标体系把描述荒漠化的各种指标如成因、发展状态或是社会指标混杂在一起。但是指标多而代表性差。不同地形如过牧、轮牧、开垦等土地情况土地的荒漠退化也是不能等价而论的。我们可以得知在开垦区、轮牧区和过牧区 3 众不同干扰条件下，优势鼠种的数量变动各不相同，不仅同一年中不同干扰类型的变动情况不同，而且同一干扰类型中不同年份的差异也较大。另一方面，对于荒漠化的治理，当前情况下并没有具体的可量化的解决方案。因此，如何解决评价和量化问题至关重要。



# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

## 二、问题的分析

对于问题一，首先分析附件中的数据是否存在异常值或缺失值，并进行数据处理。对于缺失的数据，由于附件中数据的时序关系等较弱，故我们将其去除不予考虑。因为数据的规模较小，存在偶然性的可能较大，为去除偶然性对评估的影响，考虑到样本数据为高维形式，我们采用基于共享型最近邻居相似度算法（SNN）进行异常值的筛检。然后，对数据进行归一化后，我们建立了投影寻踪模型，由于投影值可以很好地表现高维数据的分类结构特征，我们以此来获取样本的投影值作为评价值，最后采用 K-Means 算法聚类进行样本的评估。

对于问题二，我们选用元胞自动机（CA）分别对轮牧区、过牧区、开垦区进行仿真模拟，观察三个地区在未来几年内非退化、半退化、退化三种区域面积的动态变化情况，然后根据题目要求，修改元胞演化参数，分别模拟减少人为干扰和增加人工植被两种治理措施，观察此时三个地区的变化情况，并对最终结果的非退化、半退化、退化面积进行统计分析，以便能得出较好的量化治理措施。

本次建模总的流程图如图 1 所示：

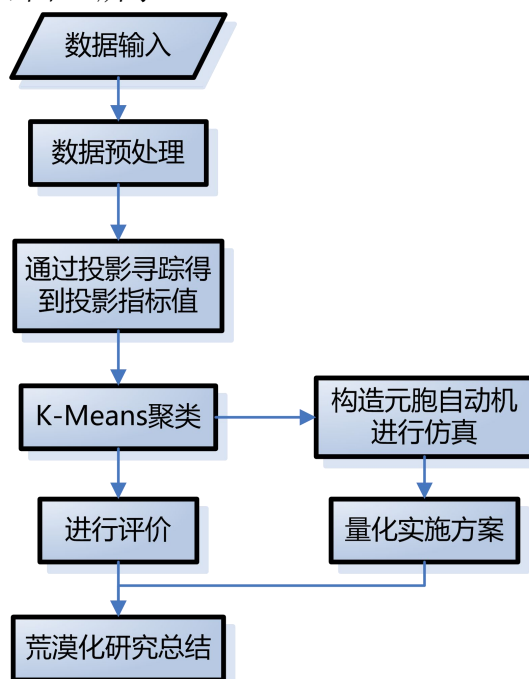


图 1 总流程图

## 三、模型的假设

- 1) 假设附件中同一月的重复项为在西北地区同一个月在不同小地域调查所得；
- 2) 由于评估生态退化的指标多种多样，指标的选取影响着评估模型的效果，假设题中所给的几项可作为生态退化的典型指标；
- 3) 假设剔除异常值之后的数据都具有统计意义，且数据来源真实可靠；
- 4) 假设附件中的小区域数据能有效地反映当地整体荒漠化的情况。

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

## 四、 符号的说明

符号	说明
$m$	评估样本量
$n$	样本维数
$D$	样本间的 SNN 相似度
$t$	检测异常评价样本的阈值
$x_{i,j}$	第 $i$ 个样本的第 $j$ 个指标值
$\sigma_z$	投影值的标准差
$\rho_z$	投影值的局部密度
$E_i$	第 $i$ 个个体被选择的概率
$W_m$	每个样本与其最近的聚类中心的距离平方和
$\lambda_1$	减少的人工干扰程度
$\lambda_2$	增加的人工植被程度

## 五、 模型的建立

### 5.1. 问题一的分析与求解

#### 5.1.1. 问题的分析

对于问题一，由于题中所给的数据有限，难以采用现有的荒漠地区生态退化的评估标准直接进行评估，故我们考虑根据数据指标自行构建评价体系来对荒漠地区生态退化的状况进行评估。因为荒漠地区生态退化的状况存在多种表现形式，在数据上表示为数据的高维形式，所以我们利用投影寻踪模型，通过最佳投影方向与评价指标的线性投影得到不同样本的投影指标值，依据投影指标值结合 K-Means 算法来进行生态退化状态的分类评估。

#### 5.1.2. 评价指标的选取

注意到附件中所提供的数据调查了白刺等植物的长势情况，白刺具有很强的固沙阻沙能力，霸王与猫头刺等具有较强的抗逆性，具有很好的饲用价值和适口性。另外，啮齿动物群落作为荒漠地区的消费者，对该地区的稳定性具有十分重要的作用。所以根据分析可知，荒漠地区生态退化的程度可以在这些调查的动植物的种群状态上很好地反映出来，故我们将附件中的植物因子与啮齿动物优势种百夹捕获率作为生态退化程度的评价指标进行以下的研究。

#### 5.1.3. 评价指标的预处理

##### 5.1.3.1. 缺失数据的处理

观察到附件二中存在着部分缺失的数据，而来自同一个月份的数据时序关系并不清楚，难以使用插值等方法进行数据的补充，故我们将附件二中存在着数据缺失的样本从总样本中去除不予考虑。

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

## 5.1.3.2. 异常数据的处理

因为采样的数据量比较少，得到的数据存在着较大的偶然性，为了避免这些偶然性对评价模型的影响，考虑到同一个月份在相同人为影响下表现的数据存在着一定相似性，我们利用了基于共享型最近邻居相似度算法（SNN）进行异常数据的筛检。

基于共享型最近邻居相似度算法（SNN）是由 Levent Ertoz 等人提出的<sup>[4]</sup>，该算法通过计算每个评价样本间的共享型最近邻居相似度来进行聚类，进而筛选出样本集中的异常样本。SNN 相似度的直观定义大致如图 2 所示：

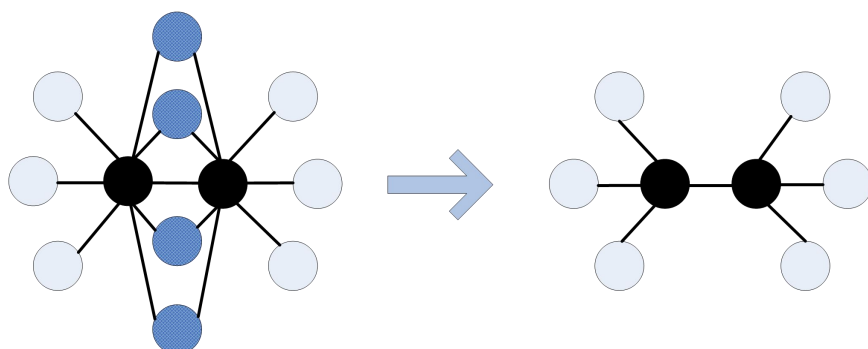


图 2 SNN 相似度直观定义图

即将计算每个样本间的欧氏距离作为相似度，进行最近  $k$  邻居的稀疏处理，以此构造最近邻居图以聚类。

算法具体的执行流程如下：

**Step1:** 计算各样本间的距离矩阵  $M$ ；

**Step2:** 依据距离矩阵建立  $k$  近邻列表  $T$ ；

**Step3:** 计算样本间的 SNN 相似度  $D$ ；

**Step4:** 检测异常评价样本：确定阈值  $t$ 。若某两个样本间的 SNN 相似度  $D > t$ ，则建立两个样本点之间的连接，将其聚为一类。最终没有和任何点聚为一类的样本作为异常样本。

用 Matlab 编程（程序见附录 1），对参数  $k$  与  $t$  的试验如表 1：

表 1 SNN 参数测试表

参数	$t=12$	$t=15$	$t=18$	$t=21$	$t=24$	$t=27$
	$n(\%)$	$n(\%)$	$n(\%)$	$n(\%)$	$n(\%)$	$n(\%)$
$k=20$	0(0.0)	9(7.4)	83(68.0)	/	/	/
$k=25$	0(0.0)	0(0.0)	2(1.6)	31(25.4)	110(90.2)	/
$k=30$	0(0.0)	0(0.0)	0(0.0)	0(0.0)	11(9.0)	55(45.1)

为了确保结果的可比性，我们最终参数的选取值为  $t = 15, k = 20$ 。相应异常点为 9 个,如表 2 异常样本表 2 所示：

表 2 异常样本

图版号	月份	重复项
35	7 月	15, 18
	10 月	16
50	7 月	5, 9, 13
	10 月	/
54	7 月	12
	10 月	3, 11



# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

至此，经过上述处理之后，将各样本按照图版号与重复项的序号大小相连，相我们得到了评价样本集  $x \in R^{m \times n}$ ，其中  $m = 113$ ， $n = 11$ 。

### 5.1.3.3. 评价指标值的归一化处理

对经过上述处理之后的评价样本集  $x$  进行讨论，由于样本数据中的动植物状况调查数据数值越高，该状态的荒漠化程度就越低，均为高优指标，同时为消除各指标值的量纲和统一各指标值的变化范围，我们采用了如下的极值归一化的方法进行了数据的预处理：

$$x_{i,j} = \frac{x_{i,j} - \min x_j}{\max x_j - \min x_j} \quad (1)$$

其中  $\max x_j$  为评价样本集  $x$  中第  $j$  项指标的最大值， $\min x_j$  为最小值。

### 5.1.4. 模型的建立

对于问题一，我们建立了投影寻踪模型，通过最佳投影方向与评价指标的线性投影得到不同样本的投影指标值，据此结合 K-Means 算法来进行生态退化状态的分类评估。具体求解流程图如图 3 求解流程图图 3 所示：

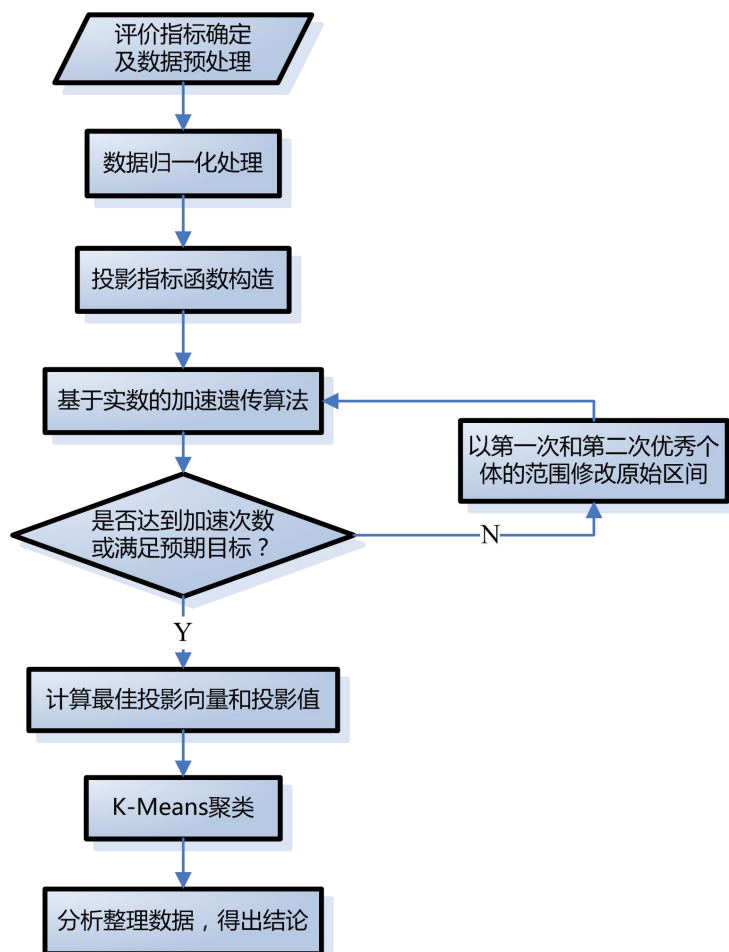


图 3 求解流程图

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

电话：0471-4969085

邮编：010021

Email: 2015@tzmcm.cn

参赛队号 #1134

投影寻踪法 (PP) <sup>[5]</sup> 属于直接由样本数据驱动的探索性数据分析方法。它把高维数据  $x \in R^{m \times n}$  通过某种组合投影到低维子空间上  $\{z(i)\}$  进行聚类, 通过分析低维空间数据点的变化特征来研究高维数据结构特征。聚类就是按照一定的规则或者标准对数据样本进行分类。投影寻踪法是以类内之间具有较大密度集而类间具有相对大的散开度为目标寻找一维投影方向, 并根据每个样本相应的综合投影值对样本进行综合评价分析的方法。

我们运用投影寻踪法在评价样本集  $x \in R^{m \times n}$  下进行讨论。其中  $x_{i,j}$  为第  $i$  个样本第  $j$  个指标值。

首先构造投影指标函数。投影寻踪法就是把  $n$  维数据  $\{x_{i,j} \mid j=1 \sim n\}$  综合成以  $a = (a_1, a_2, \dots, a_n)$  为投影方向的一维投影值  $z_i$  :

$$z_i = \sum_{j=1}^n a_j \times x_{i,j} \quad (2)$$

然后根据  $\{z_i \mid i=1 \sim m\}$  的一维散布图进行分类。上式中  $a$  为单位长度向量。在综合投影值时, 要求投影值  $z_i$  的散布特征应为: 局部投影点尽可能密集, 最好凝聚成若干个点团; 而在整体上投影点团之间尽可能散开。据此投影指标函数可构造为

$$Q(a) = \sigma_z \times \rho_z$$

式中,  $\sigma_z$  为投影值  $z_i$  的标准差,  $\rho_z$  为投影值  $z_i$  的局部密度

样本序列的投影特征值标准差计算方式如下:

$$\sigma_z = \sqrt{\frac{\sum_{i=1}^n (z_i - \bar{z})^2}{n-1}} \quad (3)$$

$\bar{z}$  为所有样本在投影方向上的投影特征值的均值, 由概率统计的知识可知,  $\sigma_z$  越大, 类与类之间的散开程度越大。

设任意两个样本  $k, l$  之间的投影值之差为

$$r_{kl} = |z_k - z_l| \quad (4)$$

则投影值  $z_i$  的局部密度  $\rho_z$  可由以下公式计算得到:

$$\begin{aligned} \rho_z &= \sum_{k=1}^m \sum_{l=1}^m (R - r_{kl}) u(R - r_{kl}) \\ &= \sum_{k=1}^m [(R - r_{k1}) u(R - r_{k1}) + (R - r_{k2}) u(R - r_{k2}) + \dots + (R - r_{km}) u(R - r_{km})] \\ &= (R - r_{11}) u(R - r_{11}) + (R - r_{12}) u(R - r_{12}) + \dots + (R - r_{1m}) u(R - r_{1m}) \\ &\quad + (R - r_{21}) u(R - r_{21}) + (R - r_{22}) u(R - r_{22}) + \dots + (R - r_{2m}) u(R - r_{2m}) \\ &\quad \dots \dots \\ &\quad + (R - r_{m1}) u(R - r_{m1}) + (R - r_{m2}) u(R - r_{m2}) + \dots + (R - r_{mm}) u(R - r_{mm}) \end{aligned} \quad (5)$$

由于  $R > 0$ , 且当  $k=l$  时有  $(R - r_{kl}) u(R - r_{kl}) = R$ , 同时  $r_{kl} = r_{lk}$ , 于是可将上式整理如下:

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

$$\begin{aligned} \rho_z = mR + 2 \sum_{k=2}^m (R - r_{1k})u(R - r_{1k}) + 2 \sum_{k=3}^m (R - r_{2k})u(R - r_{2k}) + \\ \cdots + 2 \sum_{k=n-1}^m (R - r_{m-2,m})u(R - r_{m-2,m}) + 2(R - r_{1m})u(R - r_{1m}) \end{aligned} \quad (6)$$

$r_{kl}$  愈小，则类内密度愈大；反之，则类内密度愈小。 $u(R - r_{kl})$  为单位阶跃函数，当  $R > r_{kl}$  时， $u(R - r_{kl}) = 1$ ，否则  $u(R - r_{kl}) = 0$ 。 $R$  为密度窗宽度参数，其最佳取值为：

$$R = \frac{\max(r_{kl}) + n}{2} \quad (7)$$

接着优化投影指标函数。当各指标值的样本集给定时，投影指标函数  $Q(a)$  只随投影方向  $a$  的变化而变化。可通过求解投影指标函数最大化问题来估计最大可能暴露高维数据某类特征结构的最佳投影方向：

目标函数：

$$\max Q(a) = \sigma_z \times \rho_z \quad (8)$$

约束条件：

$$\text{s. t. } \sum_{j=1}^n a_j^2 = 1 \quad (9)$$

这是一个以  $\{a_j \mid j = 1 \sim n\}$  为优化变量的复杂非线性优化问题，运用 *Lingo* 等优化软件一般难以求出其解。所以我们考虑使用用模拟生物优胜劣汰规则与群体内部染色体信息交换机制的基于实数编码的加速遗传算法（RAGA）来求解上述问题。

最后，我们利用求解所得的最佳投影方向  $a^*$  计算各个样本的投影指标值：

$$z_i^* = \sum_{j=1}^n a_j^* \times x_{i,j} \quad (10)$$

利用 K-Means 聚类算法将各个样本的投影指标值分别聚为三类，对投影指标值的含义进行分析，即可定义聚类所得类别的意义，完成对样本集的评估。

## 5.1.5. 模型的求解

### 5.1.5.1. 基于实数编码的加速遗传算法求解投影寻踪模型

遗传算法是在达尔文的进化论和孟德尔的遗传学说的理论上发展得到的，是一种根据生物进化思想寻找全局最优解的一种概率优化算法，可以利用简单的编码技术和算法机制来模拟复杂的优化过程<sup>[6]</sup>。常用的标准遗传算法（SGA）由编码、生成初始群体、杂交、适应度评估、选择、变异等步骤组成，但是存在着易收敛于局部最优值、计算量大、解的精确度不高等问题，而基于实数编码的加速遗传算法（RAGA）变标准遗传算法的二进制编码为实数编码的形式，采用加速循环的方式求解使得求解的速度更为快捷，改进了标准遗传算法的部分问题，故在本问中我们考虑使用基于实数编码的加速遗传算法求解模型<sup>[7]</sup>。

RAGA 的求解步骤如下：

**步骤一：**编码。标准遗传算法采用二进制的编码方式，在算法的执行过程中不但增加了计算量，并且增加了产生额外最优解的可能性。RAGA 采用的实数编码利用了下述

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

的线性变换： $a_j = \underline{a_j} + y_j(\overline{a_j} - \underline{a_j})$  ( $j=1\cdots n$ )

其中  $[\underline{a_j}, \overline{a_j}]$  为决策变量（投影方向分量） $a_j$  的定义域。该变换将投影方向  $a = (a_1, a_2, \dots, a_n)$  编码为  $y = (y_1, y_2, \dots, y_n)$  的染色体形式，且对任意的  $j$ ， $y_j \in [0,1]$ 。

**步骤二：**初始化群体。设初始化群体个数为  $M$ ，分别在定义域  $[\underline{a_j}, \overline{a_j}]$  内随机生成  $M$  组初代投影方向  $\{dre_{j,i}^{(0)}\} (j=1\cdots n, i=1\cdots M)$ ，将其代入步骤一中进行实数编码后得到初代编码序列  $\{y_{j,i}^{(0)}\} (j=1\cdots n, i=1\cdots M)$ ，将该序列依次代入目标函数  $Q(a)$  中求取群体中各个个体的目标函数值  $\{f_i^{(0)}\} (i=1\cdots M)$ ，依据目标函数值的大小对个体进行升序排序，选择前  $Num$  ( $Num < M$ ) 个个体作为优秀个体。

**步骤三：**适应度评价。由于目标函数值越小，表示该个体的适应度越高，再者考虑目标函数值为 0 的情况，故定义适应度函数为： $S_i = \frac{1}{f_i^{(0)} \times f_i^{(0)} + 0.001}$ 。

**步骤四：**进行选择操作产生第 1 个子代群体  $\{y_{j,i}^{(1)}\} (j=1\cdots n, i=1\cdots M)$ 。第  $i$  个个体被选择的概率为  $E_i = S_i / \sum_{i=1}^M S_i$ ，定义  $\left\{p \mid p_i = \sum_{k=1}^i E_k\right\}$ ，则  $p$  将  $[0,1]$  区间划分为  $M$  个子区间： $\{[0, p_1], (p_1, p_2], \dots, (p_M, 1]\}$ 。接着随机生成  $(M - Num)$  个随机数  $Rand \in R^{M-Num}$ ，则个体选择方式为：

$$y_{j,k}^{(1)} = y_{j,i}^{(0)} \quad Rand^k \in (p_{i-1}, p_i)$$

对于剩余的  $Num$  个个体，采用上一代中的优秀个体直接补充，即得第 1 个子代群体。

**步骤五：**进行杂交操作产生第 2 个子代群体  $\{y_{j,i}^{(2)}\} (j=1\cdots n, i=1\cdots M)$ 。根据步骤四中得到的选择概率  $E_i$  随机选择一对个体  $y_{j,k1}$  与  $y_{j,k2}$  进行杂交，因为编码方式的限制，故我们采用如下随机线性组合的方式进行杂交获得子代个体  $y_{j,i}^{(2)}$ ：

$$\begin{cases} y_{j,i}^{(2)} = r_1 \times y_{j,i}^{(0)} + (1-r_1) \times y_{j,i}^{(0)} & r_3 < 0.5 \\ y_{j,i}^{(2)} = r_2 \times y_{j,i}^{(0)} + (1-r_2) \times y_{j,i}^{(0)} & r_3 \geq 0.5 \end{cases} \quad (11)$$

其中  $r_1$ 、 $r_2$ 、 $r_3$  为在  $[0,1]$  区间内的随机数。

**步骤六：**进行变异操作产生第 3 个子代群体  $\{y_{j,i}^{(3)}\} (j=1\cdots n, i=1\cdots M)$ 。采用  $n$  个随机数以  $pm_i = 1 - E_i$  的概率来代替第  $y_{j,i}$  个个体进行变异：

$$\begin{cases} y_{j,i}^{(3)} = r_4 & r_5 < pm_i \\ y_{j,i}^{(3)} = y_{j,i}^{(0)} & r_5 \geq pm_i \end{cases} \quad (12)$$

其中  $r_4$ 、 $r_5$  为在  $[0,1]$  区间内的随机数。

**步骤七：**演化迭代。将步骤四到步骤六中得到的个体依照适应度大小升序排序后，选取前  $M$  个个体构成新的种群，代入步骤三重新进行繁殖演化。

**步骤八：**加速循环。使用第一次与第二次演化迭代所产生的优秀个体对应的变量变

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

化区间作为新的初始变化区间。算法重新进入步骤一。如此循环，直至算法达到预设的加速次数  $N$  时，结束迭代。在本问中，在如下参数值下进行求解。

表 3 参数设置表

变量名	变量取值
$M$	400
$N$	7
$Num$	20

由于遗传算法本身是一个近似求解的算法，我们循环运行了 50 次，取目标函数的最大值作为投影寻踪法最终的最优值，其对应的解为最优解，即最佳投影方向。利用 *Matlab* 编程（程序见附录 2），50 次循环重复求得的目标函数值如表 4 所示：

表 4 循环求解目标函数值表

循环次数	目标函数值	循环次数	目标函数值
1	4.097149	26	4.106976
2	3.948625	27	4.063666
3	4.339008	28	4.127591
4	3.604311	29	3.692501
5	3.689384	30	4.41522
6	4.385092	31	3.563645
7	3.630286	32	3.719907
8	3.885834	33	3.933598
9	3.757874	34	3.695213
10	3.908903	35	3.876985
11	3.616843	36	4.053974
12	4.240452	37	4.047607
13	3.632974	38	3.608216
14	4.218324	39	4.094749
15	4.143327	40	3.523594
16	4.135255	41	4.229097
17	3.840803	42	4.235249
18	3.935735	43	3.781985
19	4.000772	44	3.961725
20	4.178552	45	3.98319
21	3.989729	46	4.053084
22	3.924605	47	3.925695
23	3.533991	48	4.080858
24	4.45914	49	4.002833
25	4.045293	50	4.026621

由上可见第 24 次循环求得的目标函数值是最大的，故取第 24 次循环的解作为投影寻踪法的最佳投影方向  $a^*$ ，并依据该方向求取各样本的投影评价值。 $a^*$  与部分样本的投影评价值（所有样本的投影评价值见附录 3）如下所示：

表 5 最佳投影方向及部分样本的投影价值表

样本编号	投影评价值	样本编号	投影评价值	样本编号	投影评价值
1	1.124217	11	0.3439	21	0.350176
2	0.678949	12	0.278666	22	0.687283
3	0.942425	13	0.251685	23	0.556054
4	0.895133	14	0.508732	24	0.509966
5	0.272019	15	0.464988	25	0.246017
6	0.360018	16	0.801369	26	0.269105
7	0.514011	17	1.030295	27	0.20571
8	0.699844	18	0.491107	28	0.248378
9	0.140073	19	0.477483	29	0.209693
10	0.184047	20	0.446319	30	0.168133
最佳投影方向	(0.0761, 0.1167, 0.1735, 0.1113, 0.2068, 0.3520, 0.2208, 0.1362, 0.5483, 0.5081, 0.3811)				



# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

## 5.1.5.2. K-Means 算法构造评价体系

K-Means 算法是由 *MacQueen* 于 1967 年提出的聚类算法。该算法的基本思想为对于  $m$  个样本，寻找  $k$  个聚类中心，使得每个样本点与其最近的聚类中心的距离平方和  $W_m$

最小<sup>[8]</sup>，即： $W_m = \sum_{i=1}^m \min_{1 \leq j \leq k} |x_i - a_j|^2$ 。

在算法的执行过程中，聚类中心个数  $k$  的选取会左右算法的执行效果。在生态退化的状况表示上，我们期望将样本分为退化、半退化与未退化状态，故选择  $k = 3$  执行算法，使用投影指标值作为样本的评价标准。利用 *Matlab* 编程求解得聚类结果如下：

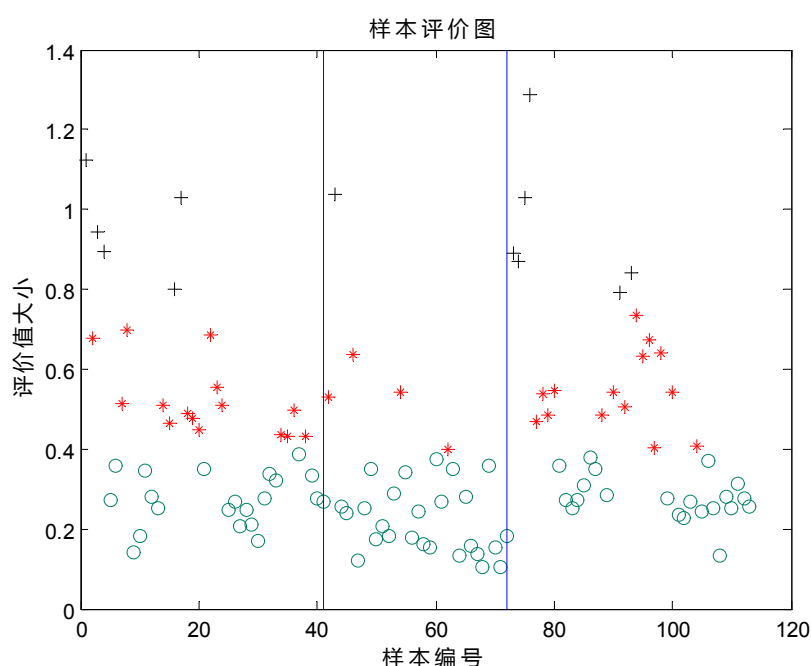


图 4 样本评价图

图中的不同图例表示不同的聚类类别，根据图中由下至上，圆圈图例表示为类别 1 (Cate1)，星号图例表示为类别 2 (Cate2)，加号图例表示为类别 3 (Cate3)。蓝色的竖线划分了不同样本的不同人类活动的样本集，其中前 41 个样本为轮牧状态下的样本集，样本编号 42 至 72 为开垦状态下的样本集，样本编号 73 至 113 为过牧状态下的样本集。

为构造出一个分类的量化标准。我们取相邻样本间的极值分界线  $range \in R^2$  作为样本的分类线：

$$\begin{cases} range_1 = \max z_j + \frac{\min z_i - \max z_j}{2} & i \in Cate1, j \in Cate2 \\ range_2 = \max z_j + \frac{\min z_i - \max z_j}{2} & i \in Cate2, j \in Cate3 \end{cases} \quad (13)$$

其中  $\min z_i$  与  $\max z_i$  分别为不同类中投影指标值的最小值与最大值。计算得分分类界线为： $range = [0.3918 \ 0.7629]$ 。即对任意的投影指标值  $z$ ，有：

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

表 6 类别范围表

所属类别	类别一	类别二	类别三
指标值大小	$0 < z \leq range_1$	$range_1 < z \leq range_2$	$range_2 < z$

分类效果如下所示：

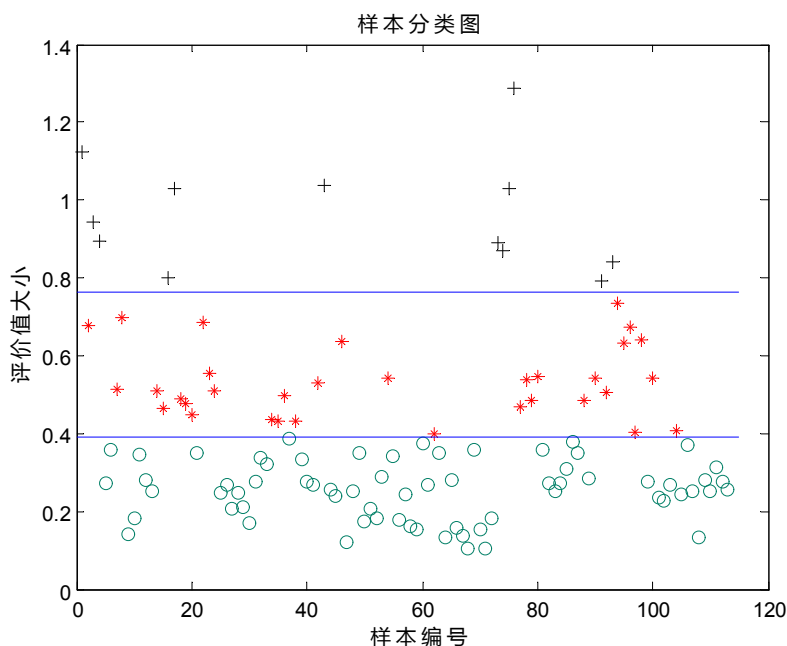


图 5 样本分类图

## 5.1.6. 求解结果的分析

求解上述模型，我们得到了样本集的分类标准，下对每个类别的含义进行分析讨论。

由于我们执行 K-Means 算法时考虑将样本分为退化、半退化与未退化状态，选择了  $k=3$  执行算法，故我们使用生态的退化、半退化与未退化状态来尝试解释分类的结果。

在投影寻踪法中，我们对评价指标集  $x \in R^{m \times n}$  进行了高优处理，而投影寻踪目标函数值  $z$  为评价指标与最佳投影方向的乘积之和，所以可以得知，最终作为分类量化指标的各个样本的目标函数值  $z_i$  也为一个高优指标，即目标函数值  $z_i$  的数值越大，该样本下展现的生态水平更好。故可对分类类别的解释定义如下：

表 7 类别定义表

所属类别	类别一	类别二	类别三
生态状态	退化	半退化	未退化

在不同人为影响下不同生态状态的样本比例统计结果如下所示：

表 8 三种状态统计结果表

生态状态	轮牧(n%)	开垦(n%)	过牧(n%)
退化	51.22	83.87	51.22
半退化	34.15	12.91	36.59
未退化	14.63	3.22	12.2

从表 8 可以看出，就总体情况来看，当地的荒漠化情况十分严重，三种人为影响下

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

均有 50% 以上的样本处于生态退化的状态，这是因为取样的地点处于西北干旱地区，当地的自然环境十分脆弱，很容易受到人为影响。

就不同的人为影响方式来看，对土地的开垦无疑是对破坏最大的，它处于退化的地区比率达到了 83.87%，为三种人为影响方式中处于退化的比率最高的，并且生态处于未退化的比率仅为 3.22%，为三种人为影响方式中处于退化的比率最低的。在本数据集下，轮牧与过牧在生态的退化比率上是相等的，在这方面并未表现出两者的差异，而观察两种人为影响方式下生态的半退化与未退化的比率可见，轮牧地区的生态处于半退化的比率要略低于过牧地区的生态处于半退化的比率，并且轮牧地区的生态处于未退化的比率要略高于过牧地区的生态处于未退化的比率，由于数据集有限，从题中提供的数据来看，轮牧对荒漠地区的负面影响要低于过牧的影响，若可以得到更多的数据集，这种影响应该会被放大而变得更加明显。

## 5.2. 问题二的分析与求解

### 5.2.1. 问题的分析

根据题目中给出的数据以及第一问的结果，轮牧、过牧、开垦这三种不同干扰方式下皆有不同程度的退化和半退化现象，因此我们需要分别进行分析其未来发展趋势，以及在减少人为干扰或补充人工植被的情况下的发展趋势。为了能直观地显示这种变化过程，我们利用元胞自动机进行仿真 (Cellular Automata, CA)，模拟其发展变化的情况。

### 5.2.2. 元胞自动机 (CA)

元胞自动机是空间和时间都离散的，物理参量只取有限数值集的物理系统的理想化模型，最早是由冯·诺依曼在研究能够自我复制的自动机时提出来的。其特点是，空间被分成离散的格子（方形、三角形或六边形等），且随时间而不断演化。元胞 (Cell) 处于若干个可能状态之一，其演化过程受周围元胞的影响，而且所有元胞的变化都是同时进行的<sup>[9]</sup>。利用元胞自动机的这些特质，我们将之前得到的三种人为干扰状态下的投影评价价值映射到二维元胞自动机中，具体的关系式如下所示：

$$Situation_{i,j} = \begin{cases} \text{退化} & , \quad 0 \leq z_{i,j} \leq range_{i,j} \\ \text{半退化} & , \quad range_{i,j} < z_{i,j} \leq range_{i,j} \\ \text{非退化} & , \quad range_{i,j} < z_{i,j} \end{cases} \quad (14)$$

由于原始数据是一维的，向多维映射时采取交叉重叠的原则进行，即

$$\begin{cases} v = randperm(n) \\ z_{i,j} = (z'_i + z'_{v(j)}) / 2 \end{cases} \quad (15)$$

对于元胞之间的邻居关系，有 VonNeumann 邻居和 Moore 邻居，如图 6 所示：

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

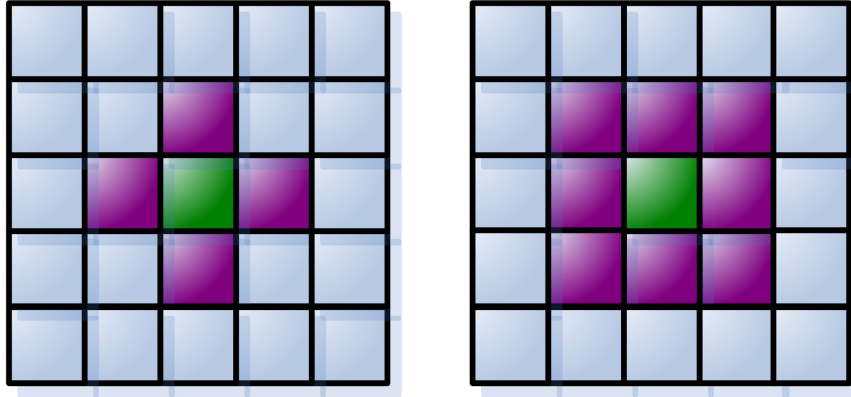


图 6 Von Neumann 邻居和 Moore 邻居

绿色为所选择的元胞，紫色为其邻居，浅蓝色为其他元胞。其中 *VonNeumann* 邻居即上、下、左、右四个方向的相邻元胞，*Moore* 邻居再次基础上增加了左上、左下、右上、右下四个相邻元胞。我们研究荒漠化问题，相邻区域之间都会有影响，因此选择 *Moore* 型邻居更加贴近实际情况。

对于元胞的演化过程，要受到自身初始值和周围相邻的 8 个元胞的影响，我们记相邻元胞的影响权值为  $weight_1$ ，自身的影响权值为  $weight_2$ 。则元胞的演化规则方程为：

$$z'_{i,j} = \sum_{x=i-1}^{i+1} \sum_{y=j-1}^{j+1} z_{x,y} \cdot f(weight, x, y, i, j) \quad (16)$$

$$f(weight, x, y, i, j) = \begin{cases} weight_1, & (x, y) \neq (i, j) \\ weight_2, & (x, y) = (i, j) \end{cases} \quad (17)$$

对于人工干扰和人工植被的影响，我们将其加入到影响权值中。对于原始的影响权值，有

$$8 \cdot weight_1' + weight_2' = 1 \quad (18)$$

而添加人工干扰和人工植被的影响后，其加和不再等于 1，而是近似接近于 1，但有一定的波动，且需要满足如下关系式：

$$\begin{cases} weight_1 = weight_1' - human + plant \\ weight_2 = weight_2' - human + plant \\ human, plant \geq 0 \end{cases} \quad (19)$$

对于人工干扰和人工植被的影响程度，采用一个系数  $\lambda$  来限制，需要满足以下关系式：

$$\begin{cases} human = human' \cdot \lambda_1 \\ plant = plant' \cdot \lambda_2 \\ 0 \leq \lambda_1 \leq 1 \leq \lambda_2 \end{cases} \quad (20)$$

为了取得较好的模拟仿真效果，并且使最终结果不至于太过极端，我们设置循环此时  $T = 30$  进行观察，此时的效果近似为几年的时间周期。

具体的仿真流程图如图 7 所示：

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

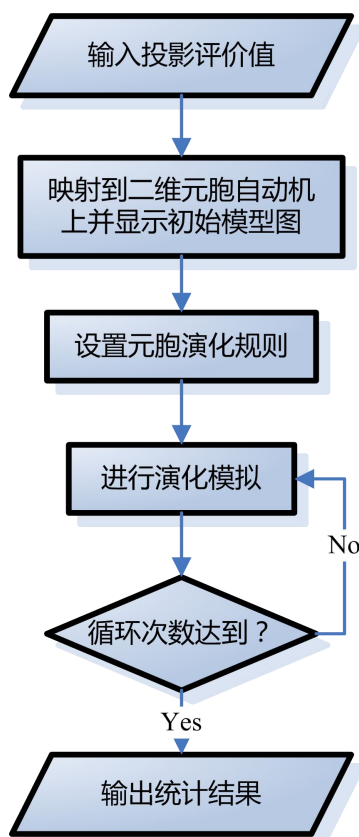


图 7 元胞自动机仿真流程图

## 5.2.3. 当前情况的模拟

根据以上模型，分别对轮牧、过牧、开垦三种人工干扰方式下的情况进行模拟，在不改变人工干扰和增加人工植被的情况下，得到以下结果：

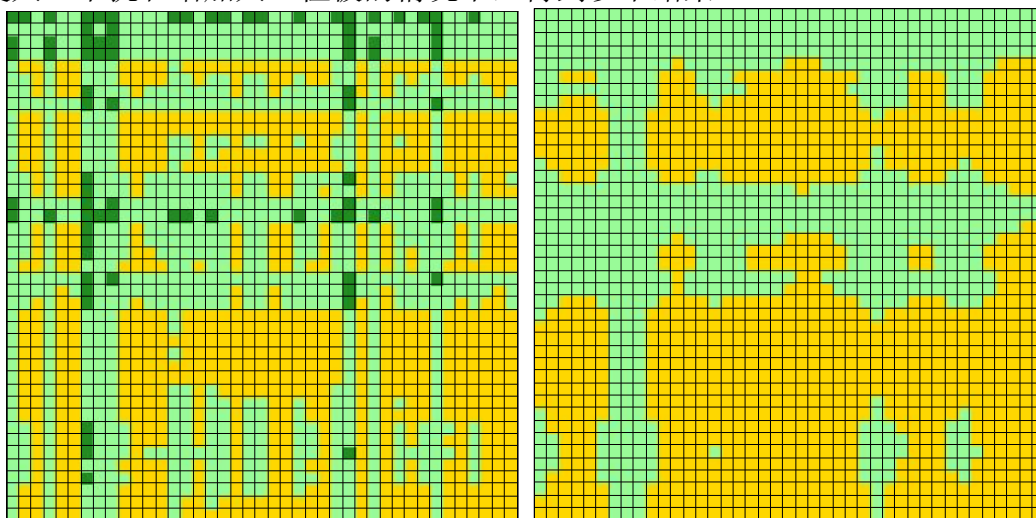


图 8 过牧初始图和过牧结果图

由图 8 可知，在过牧的情况下，初始状态时，退化现象已经有些严重，退化面积和半退化面积基本上各占一半，未退化部分分布零散，整体上呈现较为严重的荒漠化。而



# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

电话：0471-4969085

邮编：010021

Email: 2015@tzmcm.cn

参赛队号 #1134

经过模拟后，未退化的部分全部转变为半退化，荒漠化进一步加剧，已由最初的小区域恶化成大区域，荒漠化现象难以遏制。

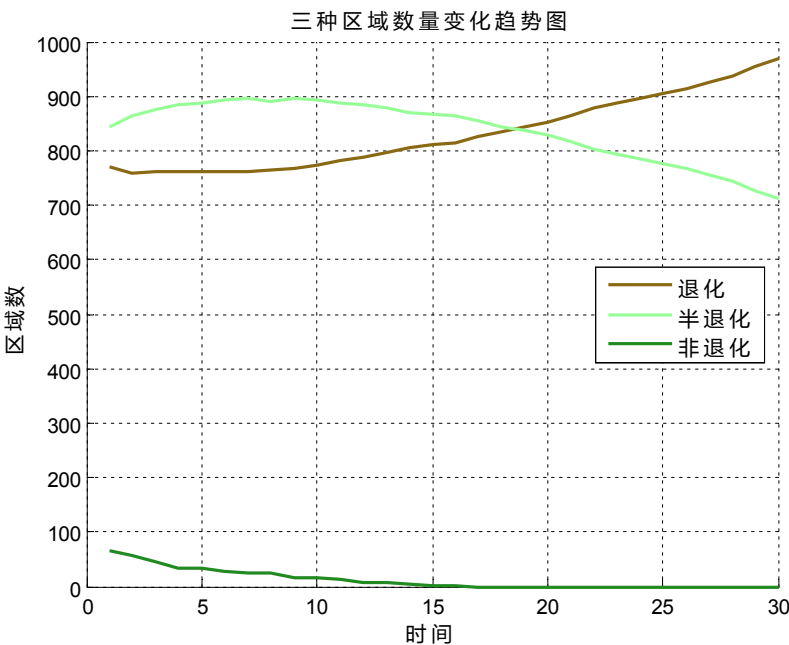


图 9 过牧结果统计图

由图 9 可清楚看到，半退化面积先增加，此时非退化面积转化为半退化面积；之后随着越来越多的半退化区域转化为退化区域，使得半退化面积逐渐减少，退化面积逐渐增加，恶化趋势逐渐加剧。

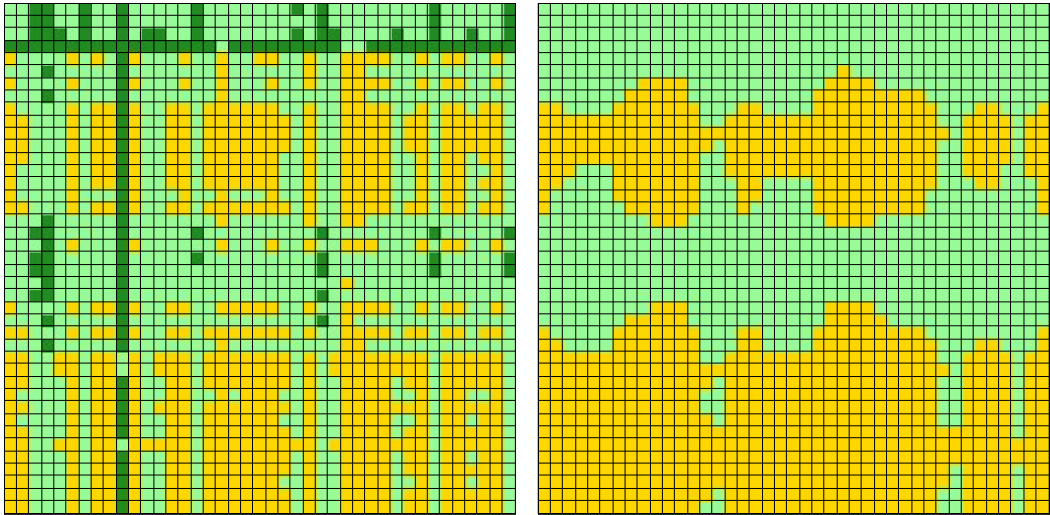


图 10 轮牧初始图和轮牧结果图

由图 10 可知，在轮牧的情况下，退化、半退化、非退化区域大致上和过牧的相同，在经过模拟后，也演变为只剩下退化和半退化区域，且退化区域逐渐蔓延成大面积区域，荒漠化情况严重。对比过牧的情况而言，轮牧的荒漠化程度相对较好，半退化面积相对较大，这和之前得到的结论是一致的。

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

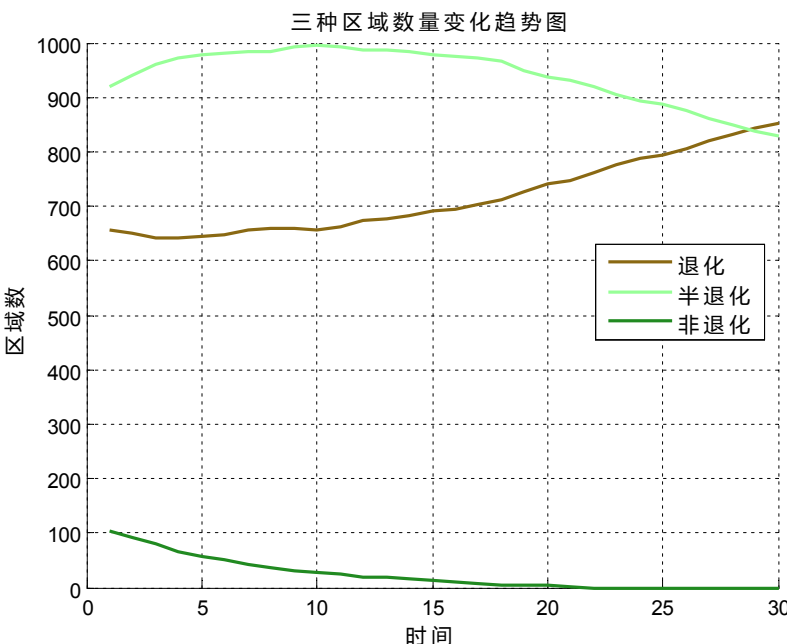


图 11 轮牧结果统计图

由图 11 可知，总体发展趋势和过牧相似，说明轮牧和过牧对生态退化的影响是相近的。相比较而言，退化和半退化的交点出现的较晚，这说明在轮牧的情况下，荒漠化趋势相对较缓，但是总体上还是在逐渐恶化，亟需人为进行治理来保持土壤。

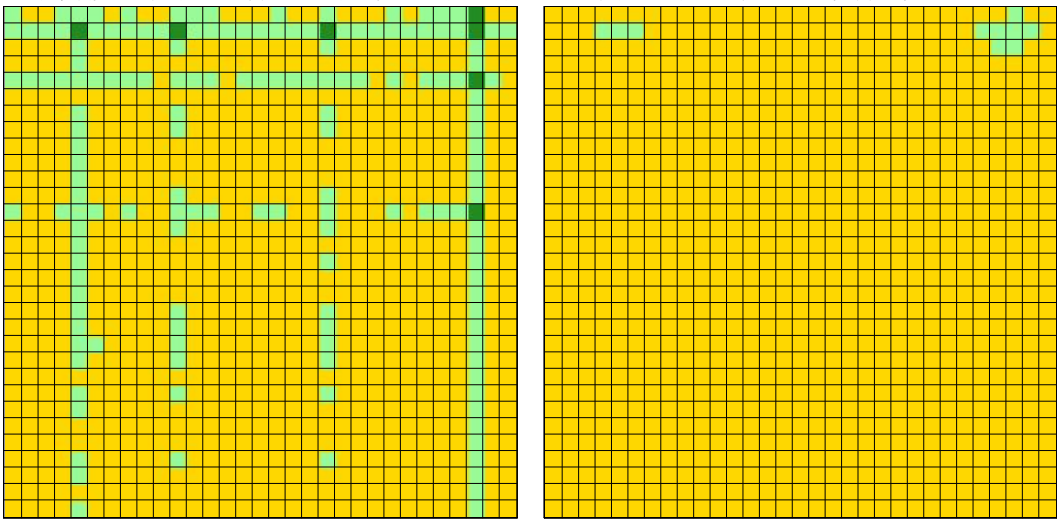


图 12 开垦初始图和开垦结果图

由图 12 可知，在开垦的情况下，初始时退化现象已经很严重了，非退化区域和半退化区域只占很少的一部分。经过模拟后，退化区域基本上占据了所有区域，全部退化只是时间问题。相对于轮牧和过牧，进行开垦时对生态造成极大的破坏，而且在相同条件下恶化更为严重，整个地区的生态退化已经难以控制。

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

电话：0471-4969085

邮编：010021

Email：2015@tzmcm.cn

参赛队号 #1134

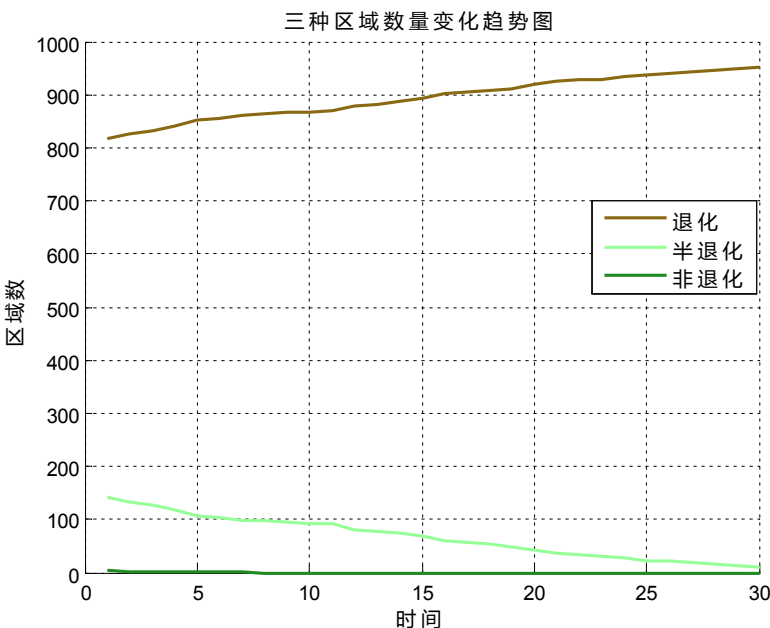


图 13 开垦结果统计图

由图 13 可知，在开垦的情况下，半退化、非退化面积从一开始就向两个极端发展，退化面积逐步占据所有区域，退化现象极为严重。

综上，当前地区退化现象严重，且经过一段时间发展后会造成极为恶劣的后果，尤其是开垦过的区域。因此，现在必须减少人为干扰或者补充人工植被来缓解当地的情况。

## 5.2.4. 采取措施补救当地环境

根据题目的要求，我们从减少人为干扰和补充人工变量两个方面进行治理研究。

### 5.2.4.1 减少人为干扰

由之前的元胞自动机模型，为减少人为干扰，减小 $\lambda_1$ 的值即可。通过对 $\lambda_1$ 值的调整，可以得到如下效果。此时 $\lambda_1 = 0.65$ ，即减少 35% 的人为干扰。

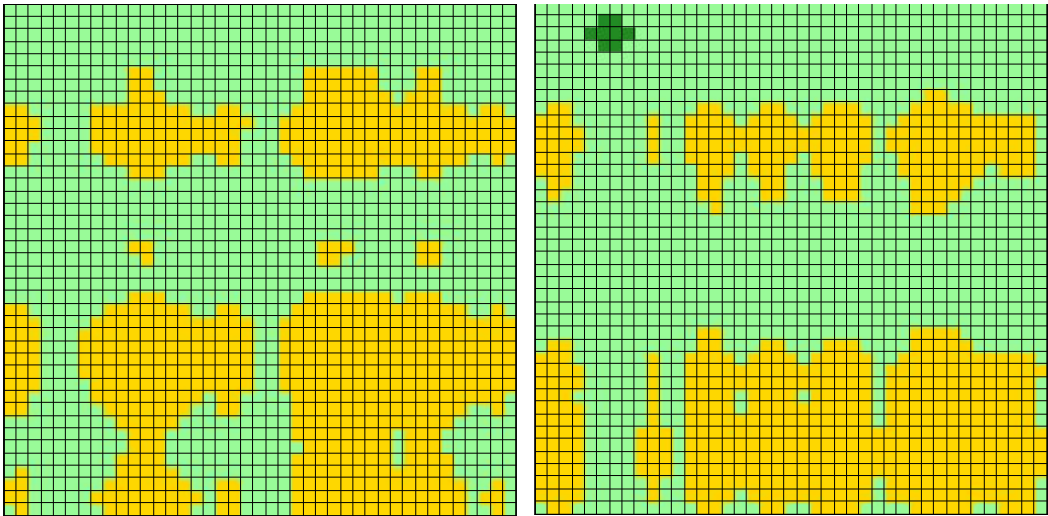


图 14 过牧结果图（左）和轮牧结果图（右）

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

由图 14 可知，此时生态退化现象并不是特别严重，非退化和半退化区域占据一定的优势，且在轮牧的条件下，还能出现非退化现象，说明生态退化现象得到了一定程度的遏制。

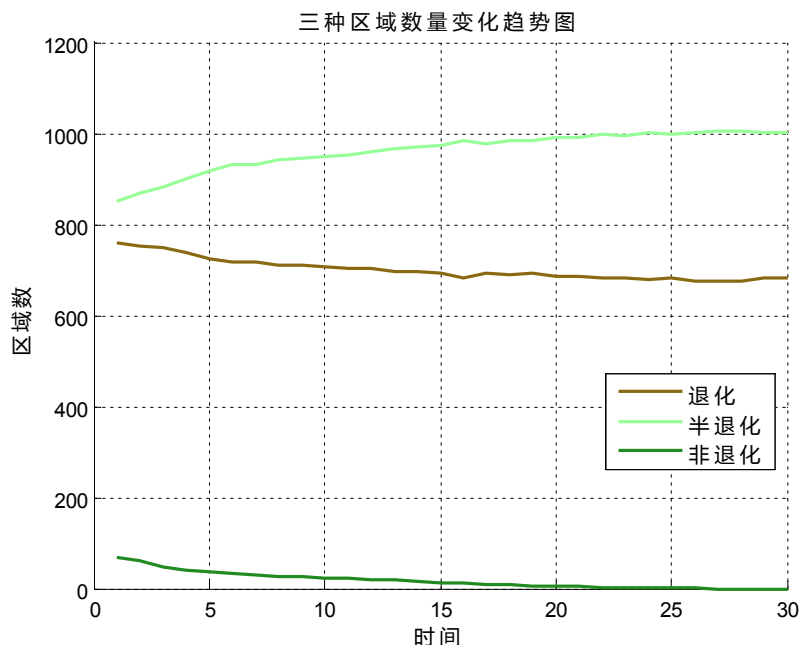


图 15 过牧结果统计图

由图 15 可知，此时退化和半退化趋势趋于稳定，而且此时半退化面积多于退化的面积，说明该地区的生态退化现象在减少 35%人为干扰的情况下已经得到了较为理想地治理，并且有进一步向好的方向发展的可能。

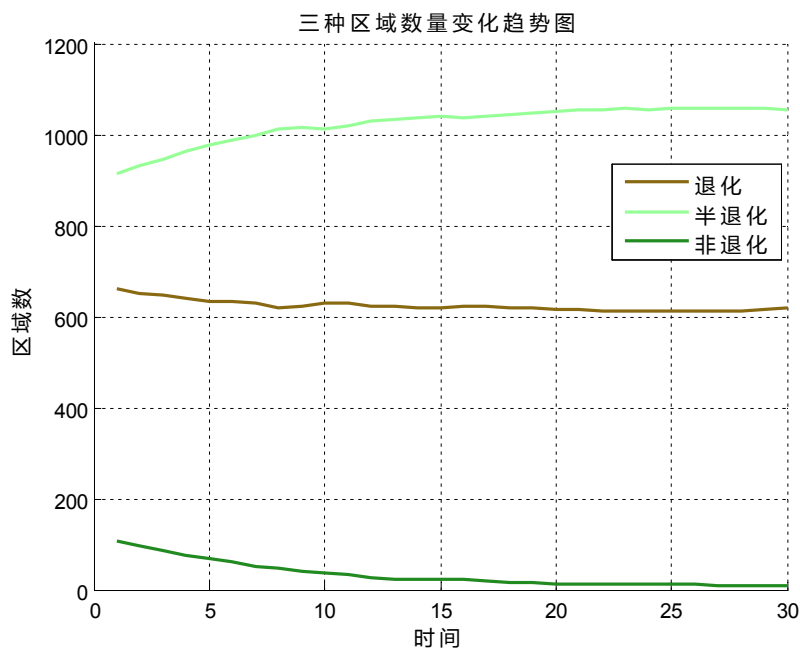


图 16 轮牧结果统计图

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

由图 16 可知，此时退化、半退化、非退化区域的面积都趋于稳定状态，且半退化和非退化面积大约是退化面积的两倍，在面积上占据较大优势。相比于过牧的情况，在减少 35%人为干扰时，轮牧的地区能得到更好的治理，并且有更好的发展前景。

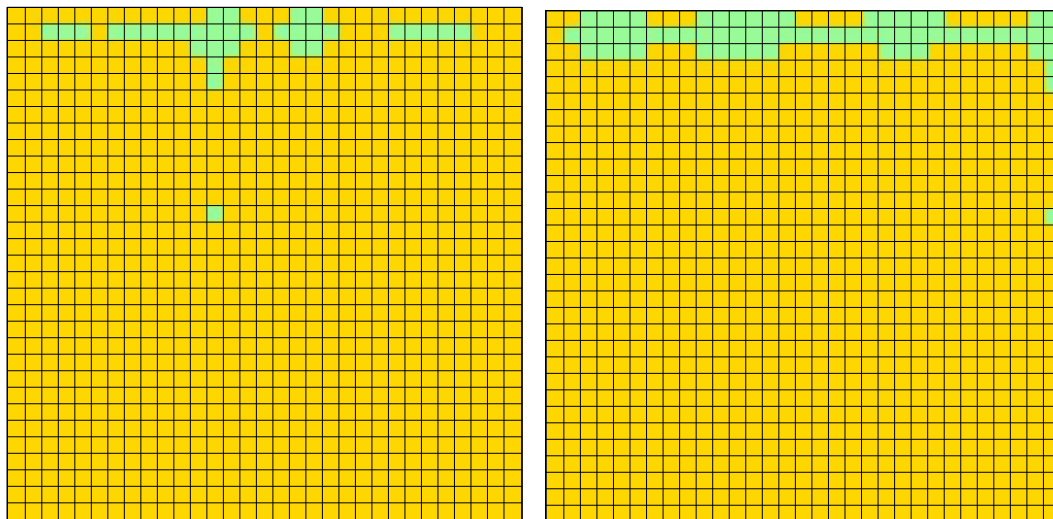


图 17 开垦结果图（左  $\lambda_1 = 0.65$ ，右  $\lambda_2 = 0.35$ ）

由图 17 可知对于开垦的区域，即使减少 35%的人为干扰也无济于事，生态退化依旧很严重；而减少 65%的人为干扰时，稍有起色，但此时退化的区域已经吞噬了所有零散的半退化和非退化区域，剩下的区域也岌岌可危。

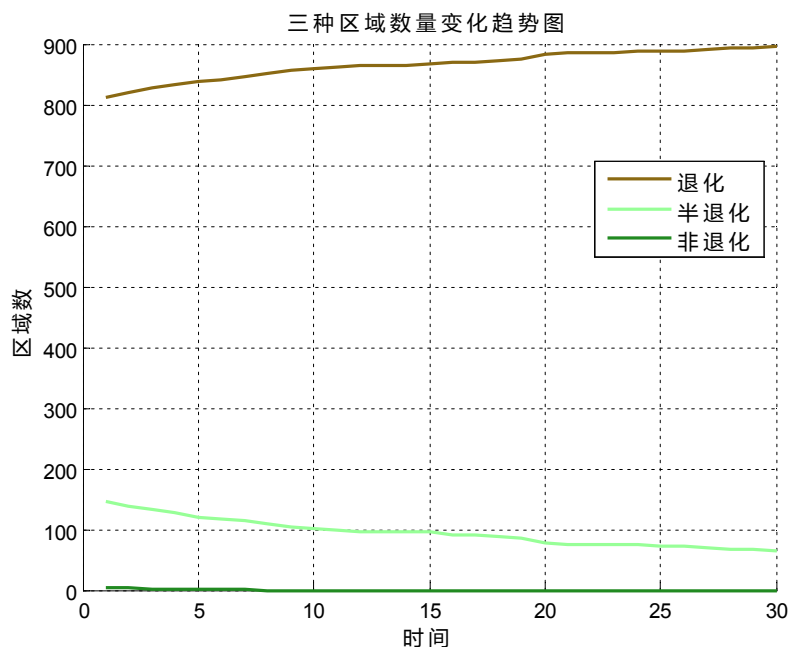


图 18 开垦结果统计图  $\lambda_1 = 0.35$

如图 18 所示，此时生态退化现象并没有太大缓解的迹象，这说明即使很大程度上减少人为干扰，对于开垦的地区来说也很难达到治理的效果。



# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

## 5.2.4.2 增加人工植被

由之前的元胞自动机模型，为减少人为干扰，减小 $\lambda_2$ 的值即可。通过对 $\lambda_2$ 值的调整，可以得到如下效果。此时 $\lambda_2 = 2$ ，即增加一倍的人工植被。

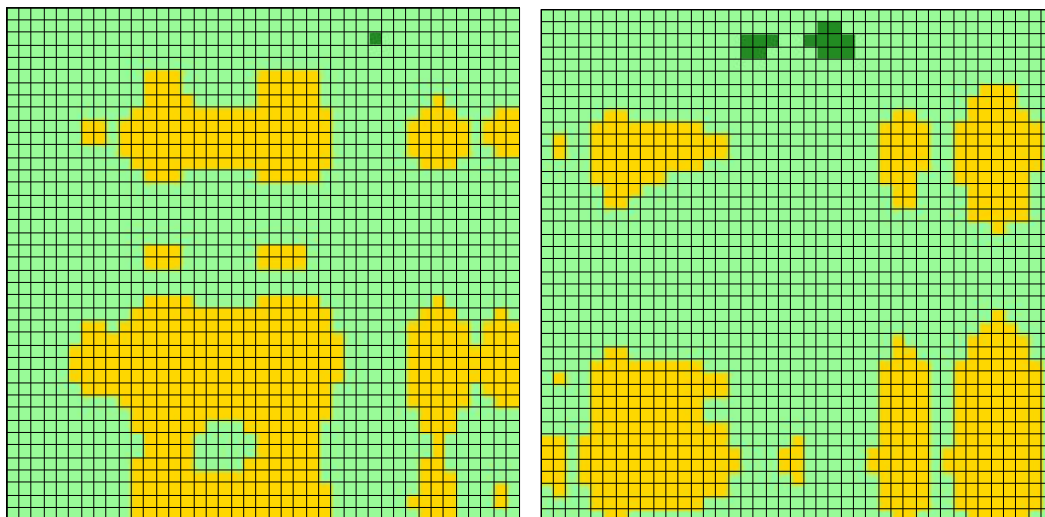


图 19 过牧结果图（左）和轮牧结果图（右）

如图 19 所示，此时退化的面积已经有了明显的收缩迹象，特别是轮牧的结果，半退化的区域占据了绝对优势，且非退化的面积聚集在一起，有进一步发展扩大的趋势。

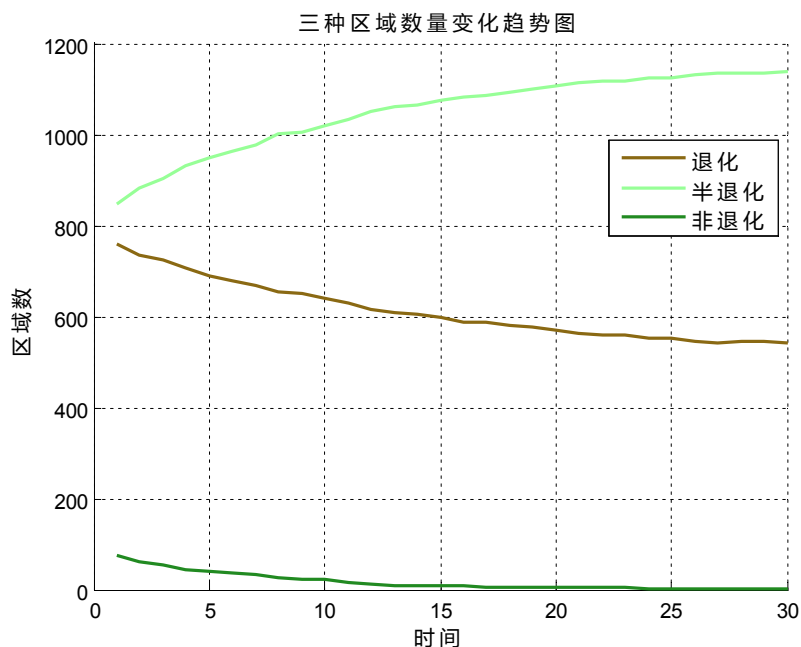


图 20 过牧结果统计图

如图 20 所示，增加人工植被后，退化的面积逐步减少，半退化的面积逐步增加，最后，半退化面积大约是退化面积的两倍。说明生态退化得到了很好的遏制，且生态有逐步恢复的趋势。

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

电话：0471-4969085

邮编：010021

Email: 2015@tzmcm.cn

参赛队号 #1134

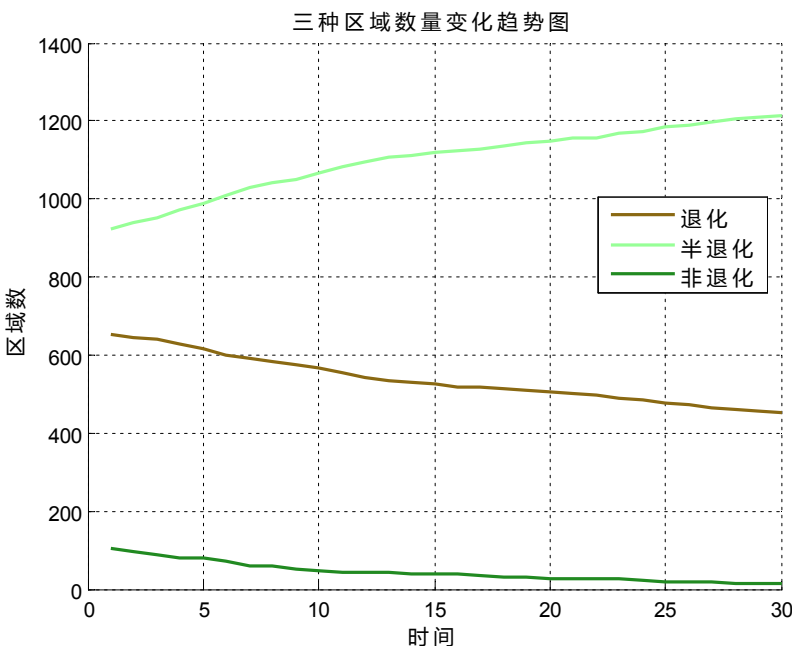


图 21 轮牧结果统计图

如图 21 所示，轮牧的情况与过牧的相比，有更好的恢复效果，半退化面积最终大约为退化面积的 3 倍。

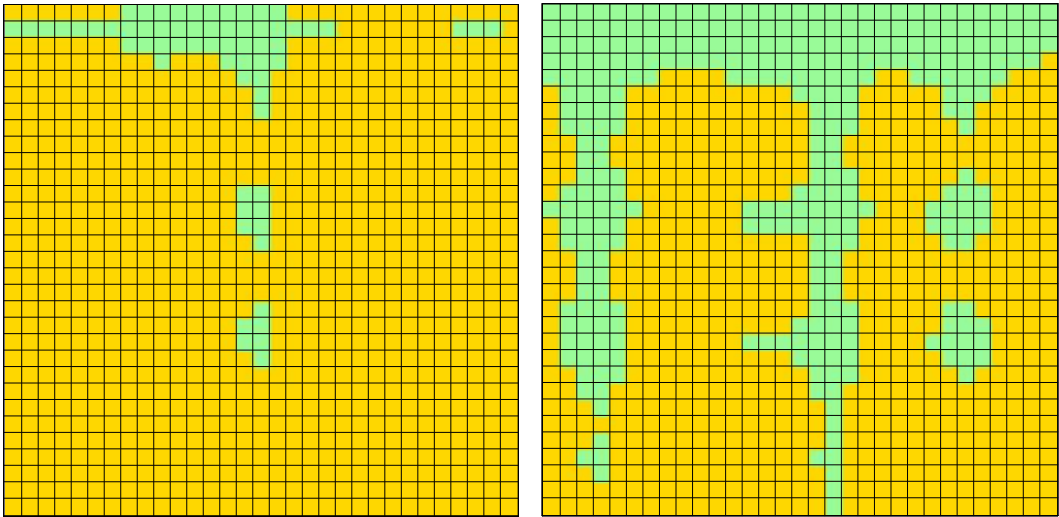


图 22 开垦结果图（左  $\lambda_2 = 2$ ，右  $\lambda_2 = 4$ ）

如图 22，在增加 1 倍人工植被的情况下，作用并不是特别明显，但是相比减少 35% 人工干扰的作用已经很好了。而在增加 3 倍人工植被的情况下，退化现象明显得到改善，半退化面积逐步增大并且呈现出良好的发展势头。

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

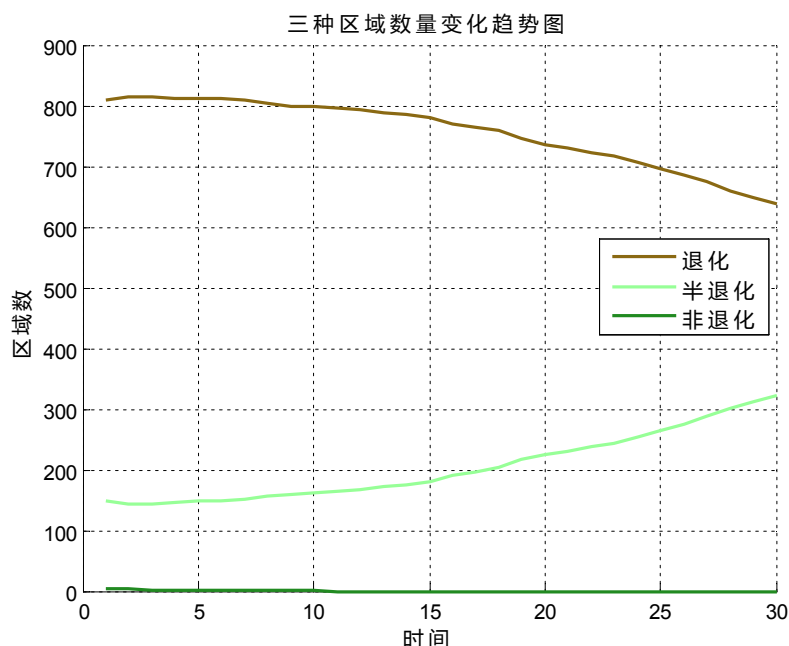


图 23 开垦结果统计图 ( $\lambda_2 = 4$ )

如图 23，此时在经过一段时间后，生态退化的地域面积开始逐步缩减，半退化的面积逐步增加。说明经过一段时间的治理之后，当地的生态退化情况有了明显的好转。另一方面，这种好的现象是在增加 3 倍人工植被的情况下才出现的，说明对于开垦的区域，想要很好地治理需要花费大量的人力和物力。

## 5.2.4.3 两种方法的总结

通过以上模型的分析可知，对于过牧和轮牧的地区来说，减少人为干扰和增加人工植被两种方法都能有效地缓解当地的生态退化情况，只需减少 35%人为干扰或者增加一倍的人工植被即可达到较为理想的效果；而对于开垦的区域来说，两种方法都不是很有效，即使减少 65%人为干扰，效果也不明显，而只有花费大量的人力物力，增加 3 倍的人工植被才能使当地的情况得到有效的改善，因此，对于开垦地区的治理任重而道远，需要寻求新的方法或者新的技术才能在花费较少的情况下治理。

## 5.2.4.4 敏感性分析

最后，针对轮牧区的情况进行敏感性分析。

表 9 敏感性分析表

human	plant	T	退化比例	半退化比例	非退化比例
0.001	0.0005	30	0.4884	0.5116	0.0000
0.001	0.001	60	0.2546	0.7448	0.0006
0.002	0.001	30	0.2927	0.7008	0.0065
0.001	0.002	15	0.1469	0.7775	0.0756

由表 9 可知，周期的长短对最终结果有显著影响，时间跨度越长，退化现象越严重，这说明当前地区的情况不容乐观，必须马上进行治理，否则后果不堪设想；人为干扰的影响较为弱，说明目前人们对于该区域的干涉已经到达了一定限度，对环境造成较大影

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

响；人工植被的影响显著，说明当前的人工绿化情况较差，没有实行太多真实有效的措施，需要加强对荒漠化的治理，提高当地居民对环境的重视情况。

## 六、模型的优缺点

在本次建模中，针对提供的未知时间序列先后的数据，我们使用了基于共享型最近邻居相似度算法（SNN）进行异常值的筛选，它的优点在于对输入样本的顺序不敏感，且具有很高的可伸缩性与处理噪音的能力。在第一问中使用的投影寻踪模型选择将高维数据投影到一个低维空间，根据投影值来分析高维数据的分类结构特征，避免了高维数据带来的计算复杂性，求解使用的基于实数编码的加速遗传算法在传统的遗传算法上进行了修改，减少了求解中计算量，使得求解的速度更为快捷。K-Means 算法作为一个成熟的聚类算法，具有坚实的数学基础与较高的泛化性能。在第二问的建模求解中使用的元胞自动机模型，在数据的反映能力上形象直观，并且作为一个动态模型，它可以十分清楚地模拟事物的变化情况。

本文的缺点在于使用 SNN 进行异常值的筛选时参数的选择存在着一定的主观性，并且题中提供的数据较少，时间序列的表现不够明确，导致结果存在一定的偶然性，若存在更多的样本数据可以分析得到更为贴近实际的结果。

## 七、模型的推广

本次建模使用的基于共享型最近邻居相似度算法（SNN）不仅可以应用于异常值的筛选，而且在聚类方面也有很好的效果。投影寻踪模型可很好地处理高维数据，在数据挖掘领域具有广泛的实用性。基于实数编码的加速遗传算法具有很强的兼容性和扩展性，可以与其他很多算法相结合进行数据的分析处理，可以模拟复杂的优化问题，在函数优化、组合优化等领域中均已得到成功的应用。而解决问题二采用的元胞自动机模型，作为一种特殊的动力系统，可用作仿真任意复杂的计算过程，可应用到天气评价，军事，产业结构、效能评估等领域。

## 八、参考文献

- [1] 丁国栋, 赵廷宁, 荒漠化评价指标体系研究现状述评, 北京林业大学学报, 26 卷第一期, 2004.
- [2] 高志海, 孙保平, 荒漠化评价研究综述, 中国沙漠, 24 卷第一期, 2004.
- [3] 胡晟男, 不同干扰下荒漠啮齿动物优势种群数量动态的研究, 2009.
- [4] 耿技, 印鉴. 改进的共享型最近邻居聚类算法[J]. 电子科技大学学报, 2006, 01: 70-72.
- [5] 赵小勇. 投影寻踪模型及其在水土资源中的应用[D]. 东北农业大学, 2006.
- [6] 金芬. 遗传算法在函数优化中的应用研究[D]. 苏州大学, 2008.
- [7] 金菊良, 杨晓华, 丁晶. 基于实数编码的加速遗传算法[J]. 四川大学学报(工程科学版), 2000, 04: 20-24.
- [8] 吴凤慧, 成颖, 郑彦宁, 潘云涛. K-means 算法研究综述[J]. 现代图书情报技术, 2011, 05: 28-35
- [9] 吕凯. 元胞自动机的研究及模型的建立[D]. 哈尔滨理工大学, 2007.

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

## 附录：

### 附录 1：SNN 除异算法

```
dis=distance(Vari_F);
[s,d]=rk(dis);
snn=rs(20,s,d);
[c,p]=cho(15,snn);
[n,m]=size(Vari_F);
Variable=zeros(n-p,m);
p1=1;
for i=1:n
    k=1;% 指示变量
    for j=1:c
        if i==c(1,j)
            k=0;
        end;
    end;
    if k==0
        continue;
    else
        Variable(p1,:)=Vari_F(i,:);
        p1=p1+1;
    end;
end;

%%
function a=distance(x2)
[n,m]=size(x2);
a=zeros(n,n);
c=0;
for i=1:n
    for j=i:n
        for k=1:m
            b(k)=x2(i,k)-x2(j,k);
        end;
        for k=1:m
            c=c+b(k)^2;
        end;
        a(i,j)=sqrt(c);
        c=0;
    end;
end;
for i=1:n
    for j=1:n
        a(j,i)=a(i,j);
    end;
end;

%%
function [a,d]=rk(dis)% a为序号矩阵, d为距离矩阵
[n,m]=size(dis);
a=zeros(n,m);
for i=1:n
    for j=1:m
        a(i,j)=j;
    end;
end;
```



# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

```
end;
end;
for i=1:n
    for k=1:m-1
        for j=1:m-k
            if dis(i,j)>dis(i,j+1)
                t=a(i,j);
                a(i,j)=a(i,j+1);
                a(i,j+1)=t;
                t=dis(i,j);
                dis(i,j)=dis(i,j+1);
                dis(i,j+1)=t;
            end;
        end;
    end;
end;
d=dis;

%%
function [c,p]=cho(t,snn)% j为相似度矩阵,c为编号, p为总个数
[n,m]=size(snn);
for i=1:n
    for j=1:m
        if i==j
            snn(i,j)=0;
        end;
    end;
end;
c=zeros(1,n);
for i=1:n
    k(i)=1;
end;
for i=1:n
    for j=1:m
        if snn(i,j)>t
            k(i)=0;
            k(j)=0;
        end;
    end;
end;
p=0;
for i=1:n
    if k(i)~=0
        c(1,p+1)=i;
        p=p+1;
    end;
end;
```

## 附录 2：基于加速遗传算法的投影寻踪模型

```
d=[];e=[];
X=Variable;
[m,n]=size(X);
for k=1:50
    x=normalize(X);
    N=400;M=10;Ci=7;n=11;DaiNo=2;ads=1;
```

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

```
[a1,b1,ee,ff]=RAGA(x,N,n,Pc,Pm,M,DaiNo,Ci,ads);
d=[d,a1];e=[b1;e];
end
[a2 b2]=max(d);
e1=e(b2,:);
ff=e1*x';
%%
function a=normalize(b)
D=b;
maxi=max(D);
mini=min(D);
[n m]=size(D);
for i=1:n
    for j=1:m
        D(i,j)=(D(i,j)-mini(j))/(maxi(j)-mini(j));
    end
end
a=D;
%%
function [a,b,mmin,mmax]=RAGA(xx,N,n,M,DaiNo,Ci,ads)
if ads==0
    ad='ascend';
else
    ad='descend';
end
mm1=zeros(1,n);mm2=ones(1,n);
Pc=0.8;Pm=0.2;
for z=1:Ci
    while 1==1
        for p=1:n
            bb(p)=unifrnd(mm1(p),mm2(p));
        end
        temp=sum(bb.^2);
        a=sqrt(bb.^2/temp);
        y=Feasibility(a);
        if y==1
            v(i,:)=a;
            break;
        end
    end
end
for s=1:DaiNo
    for i=1:N
        fv(i)=Target(xx,v(i,:));
    end
    [fv,i]=sort(fv,ad);
    v=v(i,:);
    arfa=0.05;
    q(1)=0;
    for i=2:N+1
        q(i)=q(i-1)+arfa*(1-arfa)^(i-2);
    end
    for i=1:N
        r=unifrnd(0,q(N+1));
        for j=2:N+1
            if r>q(j-1) & r<=q(j)
```

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

```
        vtemp1(i,:)=v(j-1,:);
    end
end
while 1==1
CrossNo=0;
v1=vtemp1;
for i=1:N
    r1=unifrnd(0,1);
    if r1 < Pc
        CrossNo=CrossNo+1;
        vtemp2(CrossNo,:)=v1(i,:);
        v1(i,:)=zeros(1,n);
    end
end
if CrossNo~=0 & mod(CrossNo,2)==0
    break;
elseif CrossNo==0|mod(CrossNo,2)==1
    vtemp2=[];
end
end
shengyuNo=0;
for i=1:N
    if v1(i,:)~=zeros(1,n)
        shengyuNo=shengyuNo+1;
        vtemp3(shengyuNo,:)=v1(i,:);
    end
end
for i=1:CrossNo
    r2=ceil(unifrnd(0,1)*(CrossNo-i+1));
    vtemp4(i,:)=vtemp2(r2,:);
    vtemp2(r2,:)=[];
end

for i=1:2:(CrossNo-1)
    while 1==1
        r3=unifrnd(0,1);
        v20(i,:)=r3*vtemp4(i,:)+(1-r3)*vtemp4(i+1,:);
        v20(i+1,:)=(1-r3)*vtemp4(i,:)+r3*vtemp4(i+1,:);
        temp1=sum(v20(i,:).^2);
        temp2=sum(v20(i+1,:).^2);
        v2(i,:)=sqrt(v20(i,:).^2/temp1);
        v2(i+1,:)=sqrt(v20(i+1,:).^2/temp2);
        if Feasibility(v2(i,:))==1 & Feasibility(v2(i+1,:))==1
            break;
        end
    end
end
v3=[vtemp3;v2];
while 1==1
    MutationNo=0;
    v4=v3;
    for i=1:N
        r4=unifrnd(0,1);
        if r4<Pm
            MutationNo=MutationNo+1;
```

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

```
vtemp5(MutationNo,:)=v4(i,:);
v4(i,:)=zeros(1,n);

end
end
if MutationNo~=0
    break;
end
end

shengyuNo1=0;
for i=1:N
    if v4(i,:)~=zeros(1,n)
        shengyuNo1=shengyuNo1+1;
        vtemp6(shengyuNo1,:)=v4(i,:);
    end
end

DirectionV=unifrnd(-1,1,1,n);
for i=1:MutationNo
    tempNo=0;
    while 1==1
        tempNo=tempNo+1;
        v5(i,:)=sqrt(((vtemp5(i,:)+M*DirectionV).^2)./sum((vtemp5(i,:)+M*DirectionV).^2));
        y=Feasibility(v5(i,:));
        if tempNo==200
            v5(i,:)=vtemp5(i,:);
            break;
        elseif y==1
            break;
        end
        M=unifrnd(0,M);
    end
end

vk=[v5;vtemp6];
v=vk;
end

for i=1:N;
    fv(i)=Target(xx,v(i,:));
end

[fv,i]=sort(fv,ad);
v=v(i,:);
vk=v;
vv=vk(1:20,:);
t=1:n;
mm1(t)=min(vv(:,t));
mm2(t)=max(vv(:,t));
mmin(z,:)=mm1;
mmax(z,:)=mm2;
if abs(mm1-mm2)<=0.00001
    break;
end
```

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

```
end
a=fv(1);
b=vv(1,:);
%%
function y=Feasibility(a)
b=sum(a.^2);
if abs(b-1)<=0.00001
    y=1;
else
    y=0;
end

%%
function y=Target(x,a)
[m,n]=size(x);
for i=1:m
    s1=0;
    for j=1:n
        s1=s1+a(j)*x(i,j);
    end
    z(i)=s1;
end
Sz=std(z);
R=0.1*Sz;
s3=0;
for i=1:m
    for j=1:m
        r=abs(z(i)-z(j));
        t=R-r;
        if t>=0
            u=1;
        else
            u=0;
        end
        s3=s3+t*u;
    end
end
Dz=s3;
y=Sz*Dz;
```

附录 3：投影值

样本编号	投影值	样本编号	投影值	样本编号	投影值	样本编号	投影值
1	1.124217	30	0.168133	59	0.154168	88	0.484121
2	0.678949	31	0.275642	60	0.375508	89	0.285305
3	0.942425	32	0.337324	61	0.266358	90	0.540504
4	0.895133	33	0.319949	62	0.397191	91	0.792327
5	0.272019	34	0.43681	63	0.348434	92	0.506835
6	0.360018	35	0.432714	64	0.13162	93	0.840045
7	0.514011	36	0.497991	65	0.279003	94	0.733548
8	0.699844	37	0.386391	66	0.156147	95	0.633564
9	0.140073	38	0.432989	67	0.138439	96	0.672548
10	0.184047	39	0.333406	68	0.104074	97	0.404243
11	0.3439	40	0.27469	69	0.356189	98	0.6418
12	0.278666	41	0.269714	70	0.153188	99	0.276852
13	0.251685	42	0.53188	71	0.10385	100	0.541959
14	0.508732	43	1.035693	72	0.183982	101	0.234367
15	0.464988	44	0.253968	73	0.890308	102	0.226343

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

16	0.801369	45	0.239229	74	0.870437	103	0.268626
17	1.030295	46	0.637326	75	1.028592	104	0.407101
18	0.491107	47	0.121232	76	1.285663	105	0.242186
19	0.477483	48	0.251936	77	0.469409	106	0.371792
20	0.446319	49	0.348803	78	0.539382	107	0.251851
21	0.350176	50	0.174348	79	0.483921	108	0.132972
22	0.687283	51	0.20519	80	0.546611	109	0.279616
23	0.556054	52	0.182784	81	0.358793	110	0.25275
24	0.509966	53	0.287341	82	0.272137	111	0.313109
25	0.246017	54	0.540531	83	0.249744	112	0.276401
26	0.269105	55	0.339785	84	0.271752	113	0.256058
27	0.20571	56	0.180023	85	0.308808		
28	0.248378	57	0.244387	86	0.377154		
29	0.209693	58	0.161508	87	0.351512		

## 附录 4：元胞自动机

```
clear
clc
close all;
load data;%lunmu/kaiken/guomu
data = kaiken;
T = 30;
Sit = 3;% number of situation
range = [0.3918, 0.7629];
lambda = [1 4];% rate
human = 0.001 * lambda(1);
plant = 0.0005 * lambda(2);
weight = [0.01 - human + plant, 0.92 - human + plant];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[area, Data] = Create(data);
Draw_area(area, range);
figure;
sta_data = zeros(T, Sit);
for i = 1 : T
    area = Change_area(area, weight);
    sta_data(i, :) = Statistics(area, Sit, range);
end
Draw_area(area, range);
figure;
x = 1 : T;
hold on;
plot(x, sta_data(:,1), 'Color', [0.545, 0.412, 0.078], 'LineWidth', 2);
plot(x, sta_data(:,2), 'Color', [0.596, 1, 0.596], 'LineWidth', 2);
plot(x, sta_data(:,3), 'Color', [0.133, 0.545, 0.133], 'LineWidth', 2);
xlabel('时间');
ylabel('区域数');
legend('退化', '半退化', '非退化');
title('三种区域数量变化趋势图');
grid on;

%%
function [area, Data] = Create(data)
[n, m] = size(data);
Data = zeros(n, n, m);
var = randperm(n);
for i = 1 : n
```



# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

```
for j = 1 : n
    for k = 1 : 3
        Data(i, j, k) = (data(i, k) + data(var(j), k))/2;
    end
end
end
area = Data(:, :, 1) + Data(:, :, 2) + Data(:, :, 3);
end

%%
function Draw_area(area, range)
temp = area;
[L, W] = size(temp);

Area(:, :, 1) = temp;
Area(:, :, 2) = temp;
Area(:, :, 3) = temp;

for i=1:L
    for j=1:W
        if temp(i, j) == -1
            Area(i, j, :) = [190 190 190];
        elseif temp(i, j) <= range(1)
            Area(i, j, :) = [255 215 0];
        elseif temp(i, j) <= range(2)
            Area(i, j, :) = [152 251 152];
        else
            Area(i, j, :) = [34 139 34];
        end
    end
end
Area = uint8(Area);
p = imagesc(Area);
hold on;
plot([ (0:W) ', (0:W) ']+0.5, [0,L]+0.5, 'k');
plot([0,W]+0.5, [ (0:L) ', (0:L) ']+0.5, 'k');
axis image;
set(gca, 'xtick', []);
set(gca, 'ytick', []);
end

%%
function [new_area] = Change_area(area, weight)
[L, W] = size(area);
temp = Amplify(area, mean(area(:)));
new_area = zeros(L, W);
for i = 1 : L
    for j = 1 : W
        for x = i : i + 2
            for y = j : j + 2
                if x == i+1 && y == j+1
                    new_area(i, j) = new_area(i, j) + weight(2)*temp(x, y);
                else
                    new_area(i, j) = new_area(i, j) + weight(1)*temp(x, y);
                end
            end
        end
    end
end
```

# 第八届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

参赛队号 #1134

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2015@tzmcm.cn

```
        end
    end
end
end

%%
function data = Statistics(area, Sti, range)
[L, W] = size (area);
data = zeros(Sti, 1);
for i = 1 : L
    for j = 1 : W
        if area(i, j) <= range(1)
            data(1) = data(1) + 1;
        elseif area(i, j) <= range(2)
            data(2) = data(2) + 1;
        else
            data(3) = data(3) + 1;
        end
    end
end
end

%%
function temp = Amplify(area, s)
[n] = size(area, 2);
temp = ones(n+2)*s;
temp(2:n+1, 2:n+1) = area;
end
```