**扫雷胜率与地雷密度关系的研究——数据分析报告**

**简介：**

扫雷是电脑上的一款益智游戏。游戏目标是在最短的时间内根据点击格子出现的数字找出所有非雷格子，同时避免踩雷，踩到一个雷全盘皆输。

在扫雷的地图中，一个格子有3种情况：数字、空、地雷。数字代表周围一圈8个格子的地雷个数。空格子代表周围8格不存在地雷。而这些信息游戏开始时都是不可见的，需要一点一点挖掘。此外，游戏开始点击的第一个格子必定是空格子。

本项目通过编程模拟实现自动扫雷，研究扫雷胜率与地雷密度及其他参量之间的关系。

自动扫雷算法执行流程：

1. 随机生成一个01矩阵，1表示该位置有雷。
2. 采用卷积运算，得到数字矩阵。（即标明周围有多少个雷）
3. 初始化一个矩阵know用于表示已知区域，初始全为未知。
4. 开始扫雷，选取一个空格子，作为开始的地方。
5. 对已知的格子进行广度优先遍历，采用简单逻辑判断周围的雷和数字。
6. 对所有未知格子旁边的已知格子再次进行广度优先遍历，确保当前所有可以进行简单逻辑的格子判断完毕。
7. 如果上个步骤没有开采新格子，进行复杂逻辑判断，根据多个格子的条件判断。
8. 如果进行复杂逻辑判断后还是没有开采新格子，就随机开采一个未知格子。
9. 重复进行5-8，直到开采到地雷或者全部开采完毕。

在研究中，我们引入几个概念：

1. 地雷密度：地雷数量与地图面积的比值。
2. 打开地图：地图超过80%的区域被解开。
3. 碰运气：当前情况下无法进行逻辑判断，随机挖掘一个未知格子。
4. 不靠运气取胜：全程只采用逻辑推导取得胜利，不碰运气。

此外，程序中的扫雷采用的是Win7规则。（即第一次点击的格子是空格子）

**问题一：初级中级高级到底有多难？能否完全通过逻辑推理而不碰运气取胜？**

首先，我们来看看默认的三个难度下，自动模拟程序的运行结果。其中初级难度模拟了100000次，中级模拟了50000次，高级模拟了30000次。

|  |  |
| --- | --- |
| **地雷密度——初级** | **0.123（10/81）** |
| 胜率 | 92.625% |
| 不靠运气取胜的概率 | 80.442% |
| 打开地图的概率 | 97.619% |
| 平均每局用时 | 20ms |
| 平均每局碰运气次数 | 0.31次 |

|  |  |
| --- | --- |
| **地雷密度——中级** | **0.156（40/256）** |
| 胜率 | 76.976% |
| 不靠运气取胜的概率 | 55.516% |
| 打开地图的概率 | 94.398% |
| 平均每局用时 | 68ms |
| 平均每局碰运气次数 | 0.81次 |

|  |  |
| --- | --- |
| **地雷密度——高级** | **0.206（99/480）** |
| 胜率 | 21.317% |
| 不靠运气取胜的概率 | 6.850% |
| 打开地图的概率 | 63.747% |
| 平均每局用时 | 334ms |
| 平均每局碰运气次数 | 2.82次 |

我们发现，随着地雷密度的上升，胜率和不靠运气取胜的概率出现断崖式下跌。初级在80%的情况下不需要人品，中级不碰运气取胜的概率为55%，而高级的概率就惨不忍睹了，只有不到7%。所以我们可以得出结论：初级在绝大多数情况下不需要人品，中级难度打开地图也基本不需要碰运气，而高级连打开地图都需要一些运气，不靠运气获胜的概率更是惨不忍睹。

由于打开地图的概率比胜率要高出一大截，我们可以推断：绝大多数扫雷的失败，基本都是在最后关头出现的。例如中级难度失败率为23%，但是打开80%地图的失败率只有6%不到，说明你每失败4场，就有3场是在最后20%失败的。“行百里者半九十，事将成而终止，曰为山九仞，功亏一篑。”这也是扫雷这个游戏的坑爹之处。到了最后，运气也是很重要的。

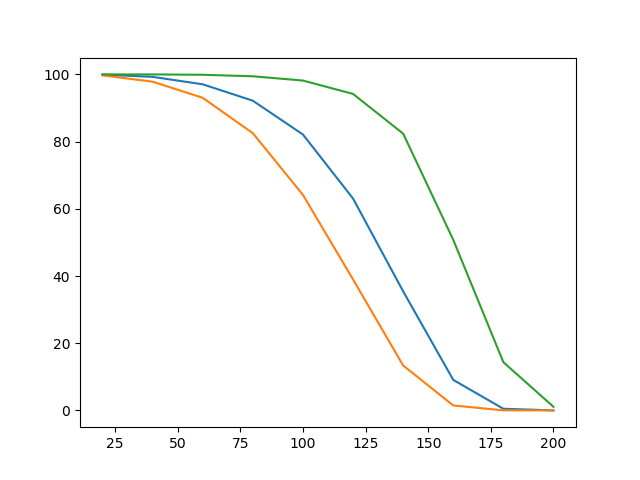
关于初级、中级、高级这三个难度到底有多难，我们在后文会进行深入讨论。

**问题二：地雷密度和胜率到底有着怎样的关系？**

我们在最大自定义地图下（30\*24）进行模拟。地雷数量从20开始，以20为公差递增到200。每种情况模拟10000次，以下是运行结果：

更详细的数据请查看文件夹内的pdf，此处仅展示曲线图。

曲线图：



橙色：不靠运气取胜的概率

蓝色：胜率

绿色：打开地图的概率

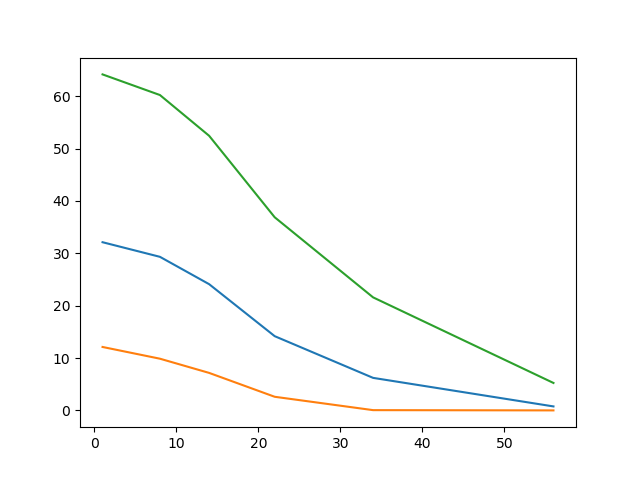
观察以上数据，我们可以发现：地雷数在90个以内的时候，胜率都在90%以上。但从90开始，胜率急剧下降，在地雷为130个左右的时候下降最快。超过175个地雷基本上没有胜算，而如果超过200个地雷，连地图都打不开。

此外我们还发现：不靠运气取胜的概率、胜率、打开地图的概率这三条曲线走势基本相同，只是开始突变时的地雷数不同。这种突变现象，在物理学中称为相变。

结论：地雷密度小于0.14或大于0.23时，胜率基本稳定（90%以上或10%以下）。而地雷密度的相变区间在0.14到0.23，这个区间内随着地雷密度上升，胜率急剧下降。

**问题三：同样的面积下，地图的形状与胜率有关吗？**

我们设定总面积为240，地雷数量为50。地图形状设为：15\*16、12\*20、10\*24、8\*30、6\*40、4\*60。每种情况模拟10000次，**取地图长与宽之差的绝对值作为横坐标**，以下是运行结果：



橙色：不靠运气取胜的概率

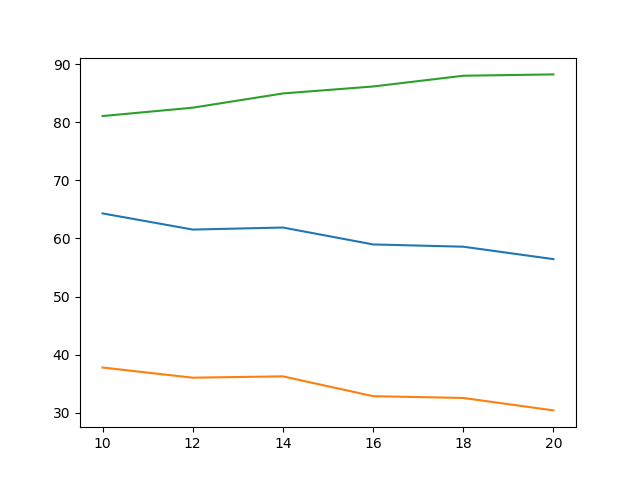
蓝色：胜率

绿色：打开地图的概率

我们发现，同样的地雷密度下，形状越接近正方形，胜率越高；形状越扁平，胜率越低。形状对胜率的作用和地雷密度类似，也存在断崖式下跌，但这个“断崖”并不明显。并且形状对胜率的影响也没有地雷密度那样显著。

**问题四：同样的地雷密度与地图形状下，扫雷的难度与总面积是否有关？**

设定地图形状为正方形，边长分别为10、12、14、16、18、20，地雷密度均为0.18，每种情况模拟10000次，**把边长作为横坐标**，以下是运行结果：



橙色：不靠运气取胜的概率

蓝色：胜率

绿色：打开地图的概率

与前两个问题不一样，打开地图的概率居然随着地图面积的增大而上升。推测是因为地图更大时，如果在一个区域无法挖掘，有更多的区域可供继续选择。由于图中胜率与地图的边长近似为线性相关，可知胜率与面积的平方根成反比。面积越大，胜率越低。但随着面积的增大，面积对胜率的影响越来越小。总的来说，面积的确对胜率有影响，但是影响不大。

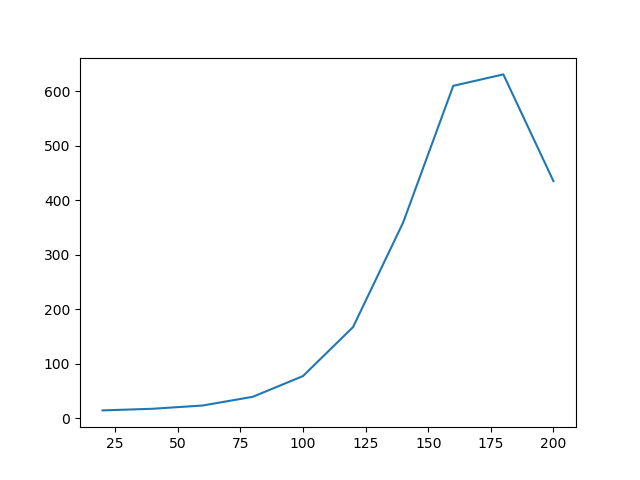
**问题五：这个算法与人类胜率的比较结果如何？得出的数据是否具有参考价值？**

从百度、知乎等论坛上得知，专业玩家高级的平均胜率大约为20%~25%左右，也就是说这个算法大概相当于一个专业玩家，可以客观反映人类玩家扫雷的结果，得到的数据有一定的参考价值。

此外，在GitHub、B站和知乎上也有着一些开源的项目。其中比较好的算法高级胜率可以达到40%以上，并且高级难度平均每局用时仅仅只有252微秒，效率与本项目中的算法拉开了很大的差距。同时，顶级玩家高级难度下的极限胜率在35%~40%之间。所以这个算法还有很大的提升空间，不过考虑到Python语言的时间常数较大，算法的效率差距可能也难以弥补。

**问题六：每场模拟耗时是否与地雷密度有关？**

对于这个问题，我们采用问题二中的时间数据，作出图像：



可以看出，在地雷密度较低（<0.22）时，随着地雷密度的增长，平均每局耗时也会出现相变式增长。但是当地雷密度超过一定程度后（>0.22），增长迅速放缓，甚至开始下降。推测是因为在地雷密度过高时，难以进行逻辑判断，游戏基本无解，只能碰运气。而碰运气在如此高的地雷密度下也无济于事，使得游戏迅速失败，耗时反而减少。

**总结：**

在本次作业中，我较好的掌握了Python语言的编程技能，能够熟练利用第三方库搭建模型解决问题。同时将数据可视化，对数据处理结果进行分析和解释，并从中得出结论。

但是，本次作业还存在以下几点不足：

1、算法的正确性和效率。由于逻辑判断算法存在缺陷。使得模拟结果只能反映一般的专业玩家的水平，可能无法进行特别复杂的逻辑判断。所以在中级和高级难度下，胜率还有提升空间。算法效率方面也有优化空间，网上优秀的算法平均时间都在微秒级别。当然，有一部分也是Python自身效率的缺陷。不过，Python的特性让我节省了很多代码量。

2、其实算法的设计方面是不必要的，可以搭建一个机器学习框架，用模型去训练。这样能够省去编写算法的步骤，当电脑训练到一定程度后，再进行模拟，理论上可以得到更加精确的结论。然而本人对这方面接触的并不多，不是很熟悉，再加上时间略微有点不足，所以没有实现这个功能。

3、难度模型的准确性有待检验。本人没有受过数学建模的培训，对于数学建模也是一知半解，建立的模型可能存在一些问题和缺陷。此外，算法的效率其实也有一部分原因，产生的数据量并不大，在本次作业中总共只进行了约400000场模拟。而其他效率更高的算法程序可以轻松模拟数百万场，得到的数据更具有一般性，更能反映模型的准确性。

**【参考资料】**

[1] 魔法小分队队长.编程实现自动扫雷，我发现了惊天秘密！[EB/OL].2017

<https://www.bilibili.com/video/BV1Xx411y7vU>

[2] 李鋆胤.利用Python编写自动扫雷程序[EB/OL].2017

<https://zhuanlan.zhihu.com/p/26079026>

<https://zhuanlan.zhihu.com/p/26090843>

[3] ztxz16.最强扫雷AI算法详解+源码分享[EB/OL].2020

<https://zhuanlan.zhihu.com/p/136791369>