**问题描述**：  
已知31\*31的网格空间内，有一信号源source（位置已知）。一个机器人（初始位置已知）从初始位置出发，根据每一步场强z(s)变化（含噪声，每走一步得到当前位置场强），寻找信号源。

要求：使用q-learning算法，通过机器人移动，找到从初始位置到达信号源的最短路径。

**问题定义**：

State：31\*31种，每个位置即一种状态

Action：当前位置的8邻域方向

相关参数：折扣因子γ= 0.8; 学习速率α = 0.1; = 0.9

Reward = z(s)（**每步更新，初始状态为0**）

Q表更新：q ( s , a ) = q ( s , a ) + α [(Reward + γ \* max(q(s’, a’) - q ( s , a )]

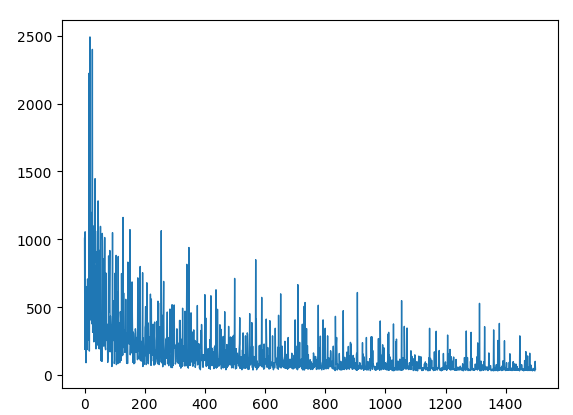
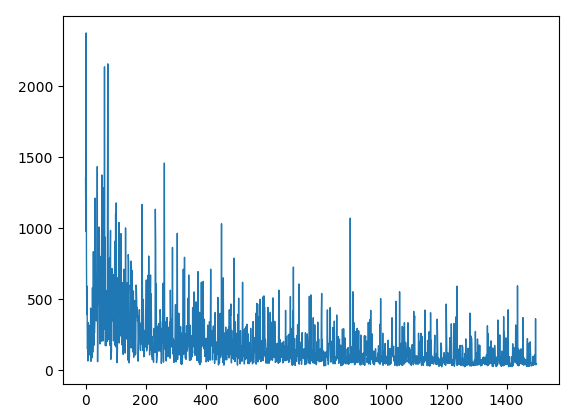
**算法描述：**

机器人从起始点出发，按-greedy（以0.9概率执行Q表最大值动作，以0.1概率随机执行动作 ）策略选择动作（下一步在当前位置的8邻域），每到达一个位置记录当前场强，以场强值作为回报，更新Q表，到达信号源周围位置即一次episode结束。

**结果：**

信号源位置一定（x轴为寻找源点次数，y轴为每次路径长度）：

最短路径为35，迭代次数为1500次



4个动作  *8个动作*

*改进：*

1. 当迭代次数足够多，可根据最终得到的Q表，从任意位置出发，都可找到到达信号源的最短路径。
2. 改变信号源位置需要重新开始更新Q表：在已知Q表的基础上，测试新的迭代次数openaigym
3. 信号源强度不断变化--因为将场强作为回报，场强更新，回报同时更新，Q表收敛后，不影响寻找最短路径。

