作业4: PacMan游戏

兰方舟 (161220059, 38633667@qq.com)
(南京大学 计算机科学与技术系, 南京210093)

1.阅读代码，阐述强化学习的方法和过程。并且回答以下问题：

- 策略模型用什么表示？该表示有何缺点？有何改进方法？
  - 策略模型为 Monte Carlo RL, 用 `QPolicy` 表示
  - 缺点: 记录特征耗费的空间过大, 计算所需时间也相当大
  - 改进: 更改模型参数和特征, 以减小时间和空间代价

- Agent.java 代码中 `SIMULATION_DEPTH`，`m_gamma` ，`m_maxPoolSize` 三个变量分别有何作用？
  - SIMULATION_DEPTH: 模拟深度, 即记录的步数, 对应强化学习课件中的T

    - T-step: $\sum_{t=1}^{T} r_t$

  - m_gamma: 衰减系数, 对应强化学习课件中的 γ

    - $\sum_{t=1}^{\infty} \gamma^t r_t$

  - m_maxPoolSize: `m_dataset` 的大小. 训练集的数据从 `m_dataset` 中取出.

- QPolicy.java 代码中，getAction 和 getActionNoExplore 两个函数有何不同？分别用在何处
  - getActionNoExplore: 选择Q值最大的Action进行下一步
  - getAction: 有 ϵ 的概率随机选择一步, $1-\epsilon$ 的概率选择Q值最大的Action进行下一步

2&3.尝试修改特征和参数，得到更好的学习性能，并报告修改的尝试和得到的结果:

- 修改特征的方法:
  - 缩减特征数量以加快训练速度. 具体的, 我们不再使用整个网格大小的 `itype` 特征, 而是仅仅使用当前avatar附近位置 (一定距离以内, 这里取曼哈顿距离不超过6) 的点的特征.
  - 加上距离特征. 由于游戏中主要避免的是被NPC吃掉, 因此计算出四个NPC和avatar的距离作为特征
  - 原本的特征数量为300个以上, 经过改进后特征数量变为100个左右, 训练速度明显得到了加快
  - 此外, 尝试只加入距离相关的四个特征, 总特征数目变为10个以内, 训练速度非常快.

- 修改参数的方法:
  - 尝试减小 γ, `SIMULATION_DEPTH` 来加快训练速度, 观察有无其它的优良效果. 具体的, 我们在这里将 γ 改为0.95, `SIMULATION_DEPTH` 改为15.

由于训练速度较慢, 我们在这里只考虑level0的情况, 其它的情况类似.

学习方法	rd1	rd2	rd3	rd4	rd5	rd6	rd7	rd8	rd9	rd10	average
原始	103	19	84	17	58	247	835	111	41	29	154
修改特征(距离和itype)	574	23	12	109	150	34	68	190	96	156	141
修改特征(仅距离)	60	25	16	109	50	234	368	190	135	215	140
修改参数	75	103	88	193	183	31	63	161	255	66	122

由此可见, 修改特征对于训练并不一定有正面的效果, 看似靠谱的特征反而很可会使效果变差. 一个很重要的原因是, 游戏所允许进行的总移动只有这1000次, 是固定的, 因此训练速度快对于结果并不会产生什么优化. 此外, 可能是强化学习发现特征的特点, 有时候可能并不需要我们人为的给定我们想当然的特征. 修改参数的也是类似的效果: 减小上述参数可以使得训练过程变快, 但是对于分数的提高并无多大作用.