

華中科技大學

研究生课程报告

姓 名	张昭骏
学 号	M202573989
系、年级	计算机科学与技术 25 级
类 别	全日制
报告科目	强化学习
日 期	2025 年 10 月 25 日

目 录

1	引言	1
1.1	Pacman 游戏	1
1.2	基于值的强化学习方法	1
1.3	报告结构	2
2	方法原理与设计	3
2.1	问题建模	3
2.2	蒙特卡洛学习方法	3
2.3	Q-Learning 方法	3
2.4	近似 Q-Learning 方法	3
3	方法实现	4
3.1	环境实现	4
3.2	蒙特卡洛学习实现	4
3.3	Q-Learning 实现	4
3.4	近似 Q-Learning 实现	4
4	实验与分析	5
5	总结	6

1 引言

在本次课程实验中，我们小组选择围绕《吃豆人》游戏环境，探索和实现了多种强化学习算法。其中我负责训练环境的搭建，并实现了蒙特卡洛学习（Monte Carlo Learning）、Q-Learning 以及 Approximate Q-Learning 三种基于值函数的强化学习方法。

1.1 Pacman 游戏

在《吃豆人》游戏中，玩家控制吃豆人（Pacman）在迷宫中上下左右移动，地图中存在固定数量的食物（food）和胶囊（capsule），以及随机移动的鬼（ghost）。当吃豆人吃掉所有食物时，游戏胜利。

游戏的基础规则可以简单概括为：吃豆人吃掉所有食物，游戏胜利；吃豆人碰到鬼，游戏失败。此外，当吃豆人吃掉胶囊后，一定时间内所有鬼会进入恐慌（scared）状态；此时吃豆人碰到鬼不会失败，而是变为吃掉鬼得分；而被吃掉的鬼会在地图中的固定位置复活，并解除恐慌。

《吃豆人》作为一个经典游戏，其状态空间大小适中，奖励结构清晰，既不过于简单也不过于复杂，适合算法的实现与测试；环境兼具确定性与随机性，鬼的随机移动增加了不确定性，考验算法的鲁棒性；此外，游戏具有长期规划特性，agent 需要在避开鬼和收集食物之间进行权衡，胶囊机制也增加了策略的多样性。

在本次实验中，我们选择《吃豆人》作为强化学习实验的环境，探索不同算法在该环境下的表现与效果。

1.2 基于值的强化学习方法

强化学习方法可以大致分为基于值的方法（value-based）、基于策略的方法（policy-based）以及结合两者的混合方法（actor-critic）。基于值的强化学习方法通过估计一个值函数（如状态值函数 $V(s)$ 或动作值函数 $Q(s, a)$ ）来指导智能体的决策，是一类经典的 model-free 强化学习方法。

在本次实验中，我依次实现了蒙特卡洛学习、Q-Learning 以及 Approximate Q-Learning 三种算法。

蒙特卡洛学习与 Q-Learning 算法是相似的，两者都通过学习动作值函数 $Q(s, a)$ 来指导决策，区别主要在于值函数的更新方式：蒙特卡洛学习基于完整回合的实际回报进行更新，而 Q-Learning 则采用时序差分（Temporal Difference, TD）方法，每一步都可以进行增量式更新，学习效率更高。通过实现这两种相对简单的方法，我加深了对强化学习基本原理和学习过程的理解。

然而，将 $Q(s, a)$ 直接作为学习目标，面临着状态空间爆炸的问题。因此，我进一步实现了 Approximate Q-Learning 算法，通过手工设计的特征函数将高维状态映射到低维特征空间，再使用线性函数逼近来估计 Q 值，从而大大降低了状态空间的复杂度，使算法能够泛化到未见过的状态，并在更复杂的环境中取得良好的表现。

1.3 报告结构

本文的组织结构如下：

第 2 章介绍三种强化学习方法的原理与设计，包括问题的 MDP 建模、蒙特卡洛学习、Q-Learning 以及 Approximate Q-Learning 的基本原理和算法设计；

第 3 章详细阐述各算法的具体实现，包括强化学习训练环境的搭建、各算法的代码实现细节；

第 4 章展示实验结果与分析，对比三种算法在不同地图规模下的性能表现；

第 5 章对本次实验进行总结，并提出未来可能的改进方向。

2 方法原理与设计

2.1 问题建模

2.1.1 Pacman 游戏环境

2.1.2 马尔可夫决策过程建模

2.2 蒙特卡洛学习方法

2.2.1 基本原理

2.2.2 算法设计

2.3 Q-Learning 方法

2.3.1 基本原理

2.3.2 算法设计

2.4 近似 Q-Learning 方法

2.4.1 函数近似原理

2.4.2 特征设计

2.4.3 算法设计

3 方法实现

3.1 环境实现

3.2 蒙特卡洛学习实现

3.3 Q-Learning 实现

3.4 近似 Q-Learning 实现

4 实验与分析

5 总结