**《项目进展汇报》**

# **一、第 [3] 周汇报（日期：[2025.7.7]-[2025.7.11]）**

1. **小组信息**

**姓名** ：唐正阳（组长）

**学号** ：20222241467

**姓名** ：陈嘉炜（组员）

**学号** ：20222241489

**姓名** ：何佩霖（组员）

**学号** ：20222241482

**姓名** ：崔忠延（组员）

**学号** ：20222241018

1. **项目概述**

**项目名称** ：基于AlphaZero 算法的五子棋自我对弈强化学习系统

**项目目标** ：实现融合AlphaZero算法与蒙特卡洛树搜索（MCTS）的五子棋 AI 系统，通过纯自玩训练构建策略-价值网络，使 AI 具备自主学习能力并达到专业水平，同时开发人机交互界面支持实战验证。

1. **本周工作内容**

**理论学习情况**

* 深入研究了MCTS的扩展和回传阶段的数学原理，重点分析了节点价值更新公式（如UCB1算法）对探索与利用的平衡作用。
* 学习了策略-价值网络训练中的数据增强技术（如棋盘旋转、镜像），以提升样本多样性。
* 探讨了自玩训练中经验回放（Experience Replay）的优化方法，包括优先级采样和批量更新策略。

**项目进展**

**MCTS完善与优化：**

* 完成了MCTS的扩展和回传阶段代码实现，确保节点回溯时价值累计逻辑正确。
* 调整了探索参数（如Cpuct），在测试棋局中验证了搜索效率的提升。
* 完成了所有内容的实现，完成了代码的全部实现。

**策略-价值网络训练：**

* 将CNN模型与MCTS结合，生成了首批自玩对弈数据（约1000局）。
* 实现了损失函数的组合（策略损失+价值损失），并完成初步训练
* 完成了五子棋以及四子棋的训练，得到了相应的模型。但是发现模型的进攻能力较强，防守能力较弱。同时模型出现了过拟合的现象。调整了训练的参数、自学习率、模型保存以及模型训练保存的条件，调整temp值，调整模型的探索深度。

**自玩训练循环：**

* 搭建了框架，支持批量数据采样和周期性模型更新。
* 修复了数据存储时的张量维度不一致问题。

**DQN算法应用**

* 同时探讨了利用DQN以及蒙特卡洛搜索来进行模型的训练，利用两个算法的自我博弈来进行巡练。
* 构建了基于DQN的模型，通过训练发现DQN模型的训练得到的模型存在过度拟合现象，在训练到10000局的情况下出现了模型出现错误的情况。

**（4）下周计划**

解决模型核心问题（过拟合与防守弱）

* 扩充数据多样性：增加随机翻转 + 噪声添加（如随机遮挡 1 个非关键落子点）的数据增强方式，将自玩数据量提升至1000局，增强模型泛化能力。
* 完成项目任务书以及相关的文档编写
* 对调整后的模型进行多局测试，对比分析参数调整前后模型在进攻、防守及泛化能力上的变化，总结优化经验。
* 整理项目进展数据，形成报告，为项目后续推进提供参考。