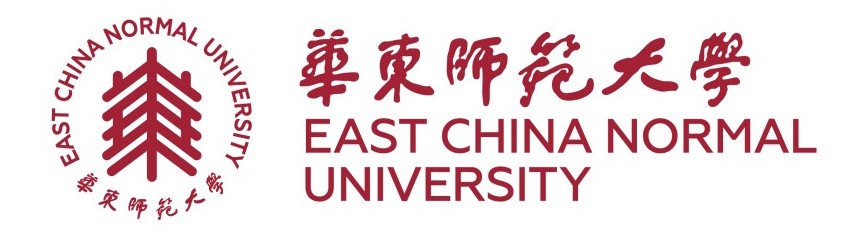
**强化学习大作业实验报告**





学 院： 数据科学与工程学院

小组成员： 赵倩仪 51255903037 李新雨 51255903028

宋晏如 51255903111 刘冬煜 51255903088

李康恬 51255903049 毛科捷 51255903071

1. **实验描述**
2. 背景知识

众包模型旨在从相对开放和快速变化的参与者群组中获得服务。众包通常使用互联网吸引和划分工作任务给参与者，并等待工作完成后将结果聚合。非盈利机构可以使用众包模型开发公共服务，比如维基百科。此外，当前也存在一些商用系统。比如，Amazon推出了Amazon MTurk系统，提供了收费的众包服务，以吸引人工解决计算机难以处理的问题。在Amazon MTurk上，请求者首先发布任务（比如：标注图片），并给出报酬（完成一张图片付费多少）。当参与者进入系统时，平台会为该参与者展示可获取的任务列表，并可查看该任务的详细信息，包括标题、描述、创建时间、结束时间等。最后，由参与者选择任务，并决定是否完成。

1. 问题描述

在众包系统中，当参与者进入平台时，涉及到任务排列的问题，即如何根据参与者选择展示的任务。为简化问题，我们假设系统只向参与者推荐一个任务。由于系统是盈利性的，因此众包系统需要同时满足参与者和请求者两方的利益：（1）参与者可以找到更多相关的、感兴趣的任务，以赚取更多的报酬；（2）请求者发布的任务可以得到更多、更高质量的回答。

与此同时，由于参与者和发布者是动态变化的，因此，系统在分配任务时应能考虑并处理这种动态性。

1. 具体任务

（1）如何将强化学习技术应用到众包任务推荐中，并最大化**参与者**的利益？

（2）如何将强化学习技术应用到众包任务推荐中，并最大化**请求者**的利益？

1. 实验要求

给出具体的实验流程和设计细节，并给出最终的实验结果。

要求同时实现并对比三种类别的方法：(1)**Value-based methods**; (2) **Policy-based methods**; (3) **Actor-critic methods**。

最多六人组队完成，于7月14日中午12点前提交实验报告至指定邮箱。

1. 实验数据

附件包括了sample\_read\_data.py文件用来读取数据（可做参考），worker\_quality.csv文件包含了worker的quality属性，project\_list.csv文件是project\_id + project\_answer\_num，project和entry文件夹分别包含了project和worker的信息。为简化问题，可直接参考sample\_read\_data.py内容读取数据。在获取了数据后，自行划分训练集、验证集和测试集，并在实验报告中说明细节。

1. 提示

（1）如何针对参与者和请求者分别定义状态、行为、奖励等？

（2）如何定义当前状态和下一步状态？

（3）如何构造Q函数？

（4）仔细分析数据，思考如何构造特征。

（5）对任务创建时间、到期时间、以及worker到达时间进行排序，模拟worker的兴趣和完成情况。

1. **实验分析**

【实验流程与设计细节】

1. 数据分析）
2. 任务分析）
3. value-based）
4. policy-based）
5. actor-critic）
6. **实验内容**

【实验流程与设计细节】（贴code，步骤说明）

1. 数据处理）
2. value-based）
3. policy-based）
4. actor-critic）
5. **实验结果**
6. 数据预处理【仪】

（数据预处理截图）

（分析）

1. Value-based methods【雨】

（该方法结果截图）

（分析）

1. Policy-based methods【如】

（该方法结果截图）

（分析）

1. Actor-critic methods【捷】

（该方法结果截图）

（分析）

1. **实验总结**

**附：**

提示解答【煜】

1、如何针对参与者和请求者分别定义状态、行为、奖励等？

A：

2、如何定义当前状态和下一步状态？

A：

3、如何构造Q函数？

A：

4、仔细分析数据，思考如何构造特征。

A：

5、对任务创建时间、到期时间、以及worker到达时间进行排序，模拟worker的兴趣和完成情况。

A：