毕设论文答辩

各位评委老师好，学长同学们好，我是来自自卓1601的薛博阳，我的指导老师是罗云峰老师，我的毕设答辩题目是《基于虚拟遗憾最小化的两人有限注德州扑克》，我将从发展背景，理论知识，实验验证以及总结展望四个方面介绍我本次的答辩内容。

首先是背景部分，这部分先简单介绍一下德州扑克发展的几个重要事件，德州扑克是机器博弈的一项重要研究内容。2006年，国际人工智能会议开始举办年度机器博弈大赛，吸引了一大批学者投入到人工智能与博弈论的研究，是现代机器博弈发展的温床。2007年，加拿大阿尔伯特大学的计算机扑克研究小组首次提出结合纳什均衡将遗憾最小化技术应用到德州扑克上，实现了德州扑克研究的飞跃式发展；2015年，再次提出改进的CFR+算法，至此宣告两人有限注德州扑克已取得弱拆解，人类已经可以训练出具备较高智能的非完备信息机器博弈体。德州扑克是典型的非完备信息博弈，游戏中每个玩家对场上的全部信息并不完全知晓，还具有随机发牌和大规模等特性，这些都使得其算法研究十分具有挑战，同时由于德州扑克风靡全球广受欢迎，完整的规则体系使得其研究具有很强的现实参考意义，而作为形式上最简单的两人有限注德州扑克，一直都是学者们的研究热点。

第二部分介绍相关的理论知识。首先引出机器博弈的表示方法，通常使用正则博弈和扩展式博弈两种表示方法，德扑是一项扩展式博弈，它的特点是玩家轮流做决策，一般采用博弈树的形式表示，这种游戏往往信息集较大，复杂度较高。

接下来是纳什均衡，纳什均衡是针对博弈游戏提出的概念，可理解为一种博弈策略，任何玩家单方面违背纳什均衡策略后，他获得的期望收益一定不会高于某一常数。从长远角度看，纳什均衡一定是最优的策略。通常理想情况下，对于零和博弈都求解0-纳什均衡，但在实际应用如本课题的德州扑克，往往通过反复迭代逼近的方式求得其 纳什均衡解。

接下来介绍虚拟遗憾最小化算法，首先遗憾是指没选取某动作产生的收益与采取当前动作的收益之差。该算法本质上是一种反复迭代逼近纳什均衡的优化算法，大致思路是将整体的遗憾分布到独立的信息子集中，采用深度优先遍历，计算每个节点的收益值和遗憾值并回溯至上一级节点，再根据返回的收益结合遗憾匹配法选择最优策略。

由于德州扑克游戏状态规模较大，需要引入合适的抽象方法来降低时间和空间的复杂度。通常分为卡牌抽象和行为抽象。行为抽象通过对玩家动作以及加注回合进行限制减少博弈树的节点，卡牌抽象最常用的是9-bucketing策略，对两张手牌进行牌力评估然后分到9个不同的区间，同一区间的手牌组合在CFR遍历过程中可以视作相同的状态节点，如左图是对两张手牌进行9-bucketing分区的示意图以及矩阵表示。

第三部分是实验验证，首先介绍我在本课题的程序框架，分为游戏平台，博弈策略，文件存储三部分，先搭建两人有限注德扑的游戏平台，再引入CFR算法进行纳什均衡策略的训练，最后将节点的遗憾值策略值等信息存储在策略文件中进行更新。

德扑的人人对战平台使用面向对象编程，主要有游戏平台，玩家，卡牌，荷官，动作，牌级等几类，游戏平台类创建对局游戏，通过对其他类的调用，实现完整的两人有限注德州扑克对局的模拟过程。

两人有限注德扑的CFR算法流程图如图，右边是一棵博弈树。算法实现的步骤为每选取一个动作就创建了一个新的节点，采取深度优先遍历，检测到叶子节点就将收益值回溯上一级，计算不同动作的遗憾值和收益值接着选择策略，再将平均收益返回，通过计算各个独立信息集的遗憾值决定整体的策略选择。

下面是我在本课题中对原有9-bucketing策略提出的改进方法，原抽象方法只关心两张手牌的牌力，本方法考虑不同阶段公共牌增加所带来的局势变化，对当前手牌与所有公共牌组合选择最大牌力进行评估分区，所以相比之前，该方法除了在翻牌前阶段仅知道两张手牌信息的情况下采用原9-bucketing分区，在已知公共牌信息的其他阶段，评估对象都是五张牌，因此一棵博弈树上的牌力bucketing也会发生变化，从而实现动态调整策略，更符合现实场景，同时也不会引起存储空间的增加。由于缺乏先验统计知识所以本方法采取均匀分布，五张牌共有2598960种组合，均分后每个bucketing约有27w种组合，如右图所示，左边是计算的各牌级手牌组合情况，右边是改进的9-bucketing分区方式。

接下来是数据及抽象动作，为了便于存储，我将所有动作序列以数字表示，如左图上是一段节选的动作序列，同时采取最多四轮加注，在一轮中无人加注就进入下一阶段等行为抽象方法；由于叫注和跟注的筹码相同所以视为相同动作，每个信息节点文件除了包含动作序列还会存储牌力bucketing区间以及可选取动作的遗憾值和策略值等信息。规定下注筹码为1，加注为2，生成的一棵博弈树共有119120个叶子节点和79798个非叶子节点。

本课题一大难点任务是训练，受疫情影响无法到实验室训练，因此在网上租用百度云服务器。对比后选择8核CPU,16GB,2.6GHz主频的计算型C4服务器；使用PyCharm Professional 远程调试。输入服务器的公网IP以及端口后等信息远程连接，同步服务器与本地数据后就可以训练，采用多进程运行方式。训练中需要评估算法收敛性来判断是否逼近纳什均衡，使用平均遗憾值及其变化率代替可利用度表示算法收敛性，如右图为训练500次的算法收敛示意图；训练中每遍历一次博弈树产生策略文件326.9MB，消耗时间为13.2s。结合9-bucketing抽象生成的策略文件共有5.75GB，训练500次耗时约2h。

最后是实验分析与不足。首先分别与纯策略为叫注和加注进行测试，结果如左图所示，上图是改进的9-bucketing策略与纯策略的对局表现，可以看出训练的策略收益为正，是一种逼近的纳什均衡策略，由于加注类似于一种诈唬手段，所以训练的策略与加注纯策略对局胜率偏低。下图是纯策略为叫注时两种抽象方法在200局游戏中的净胜筹码，分析结果发现原抽象方法相对保守，弃牌次数较多，改进的9-bucketing方法策略选取相对灵活。对于本课题，分析后认为仍存在以下三点不足，首先是训练次数不够导致改进的抽象方法对于某些游戏状态未能涵盖，接着是抽象方法过于简化使得模型太简单达不到预期，最后是为防止被对手建模采取轮盘赌转法需要付出一定的额外代价。

最后一部分是总结展望，我的各部分工作如下，先搭建德州扑克游戏平台；再结合CFR算法训练纳什均衡策略，然后使用改进抽象方法优化模型，最后租用云服务器进行测试分析，并得到了预期结论。封装好的完整应用程序及代码已上传到GitHub，下载并配置环境后即可运行，可选择人人模拟或者人机对局两种模式，人机对局的智能机器是本次课题训练的基于CFR算法的机器博弈体。另外毕设期间所写的相关CSDN博文为防止查重还未发表。

这次课题让我对机器博弈有了一个初步的认识，最主要有以下两点心得，通过毕设发现目前最大的限制仍然是硬件条件，主要是对大规模博弈的无能为力，使用抽象方法虽然降低了复杂度，但不可避免的带来信息损失，相信随着硬件计算能力的发展，已有的模型会得到改善。第二就是对统计模型缺陷的思考，一切的博弈行为均是人类社会映射在数学空间的抽象模型，如果脱离了实际情况，单纯的利用统计方法研究模型内部的数学关系，是达不到最佳的效果的，因此我认为未来应该回归研究本体，将统计模型与规则模型进行结合。如我本次答辩所提出的改进9-bucketing抽象方法就是一种不拘泥于模型限制，使得其训练过程更贴近实际游戏对局场景。