

华中科技大学

本科生毕业设计（论文）开题报告

基于虚拟遗憾最小化的两人有限注德州扑克

院 系 人工智能与自动化学院

专业班级 自卓 1601

姓 名 薛博阳

学 号 U201614481

指导教师 罗云峰

2020 年 2 月

开题报告填写要求

一、 开题报告主要内容：

1. 课题来源、目的、意义。
2. 国内外研究现状及发展趋势。
3. 预计达到的目标、关键理论和技术、主要研究内容、完成课题的方案及主要措施。
4. 课题研究进度安排。
5. 主要参考文献。

二、 报告内容用小四号宋体字编辑，采用 A4 号纸双面打印，封面与封底采用浅蓝色封面纸（卡纸）打印。要求内容明确，语句通顺。

三、 指导教师评语、教研室（系、所）或开题报告答辩小组审核意见用蓝、黑钢笔手写或小四号宋体字编辑，签名必须手写。

四、 理、工、医类要求字数在 3000 字左右，文、管类要求字数在 2000 字左右。

五、 开题报告应在第八学期第二周之前完成。

一、课题来源、目的、意义

1.1 课题来源

自人类文明诞生之日起，就有了各种各样的博弈游戏。游戏是人类早期集益智与娱乐为一体的活动，传说四千年前就有了围棋。几个世纪以来，人们创造出不计其数的各类游戏，比如象棋、国际象棋、跳棋、扑克、麻将、桌游等。半个多世纪前，电子计算机诞生，从此各种博弈游戏焕发了新貌，人类进入机器博弈的时代。

随着信息技术的迅猛发展，近几年来人工智能一直活跃在人类的舞台上，在诸多领域成就斐然，博弈论就是人工智能重要的分支之一。在计算机的早期发展中，游戏就与博弈论和人工智能紧密相连，创建具有高智能的机器博弈程序一直是计算机发展史上的一个重要目标。早在第一台计算机诞生之前，计算机之父冯·诺依曼就提出极大极小值定理。紧接着人工智能之父图灵设计了一个下棋的程序，信息论创始人香农分别为程序实现提供了硬件基础，这一程序的诞生验证了计算机科学与人工智能早期的学术观点。在半个多世纪以来，机器博弈作为新算法思想的试验平台，推动人类科技在这方面取得的成就也是人工智能发展的重要里程碑。1994 年计算机程序 Chinook 在跳棋上击败人类顶级玩家，成为第一个赢得人类冠军的机器算法；1997 年 IBM 开发的深蓝（Deep Blue）在国际象棋上击败 Kasparov，机器博弈领域发展如火如荼。

然而，牌类游戏与棋类游戏不同。国际象棋、中国象棋和围棋等都是“完备信息”游戏，也就是说，所有玩家在游戏中做决策之前知晓游戏过程中已发生的情况。AI 攻克这些游戏的难点，主要在于游戏过程的决策点数量，这决定了需要的

计算量。然而，扑克是一种包含很多隐藏信息的“非完备信息”游戏。玩家对场上已发生的情况并不是完全知道，比如对手手中的牌，更不知道对手如何猜测自己的手牌。因此，虽然一局德州扑克的决策点数量远远少于一盘围棋，但是不确定性的加入，使得每个决策点上，玩家都要全盘进行推理，计算量难以想象。这些情况对计算机的理论算法提出更高的要求，迄今为止，非完备信息机器博弈并没有像完备信息机器博弈那样得到完美的解决。如果考虑心理层面的博弈，有别于机器，人类可以“诈唬”来虚张声势，这被人类看作是智商和情商的完美结合。

机器博弈是人工智能的重要分支，目前随着 AlphaGo 解决了最复杂的完备信息博弈问题——围棋，机器博弈的研究中心已转到研究成果并不突出的非完备信息上。2007 年 Albert 大学的 Martin 等人提出了基于虚拟遗憾最小化（CFR）的概念来解决大型博弈游戏。这已成为解决非完备信息的一种思路，本课题通过实现 CFR 算法，为以后的非完备信息机器博弈研究做铺垫。2015 年 Albert 大学 Bowling 等人在《Science》上发表文章，对 CFR 进行改进，使得两人有限注德州扑克取得了突破性的进展，鼓舞了无数学者。现如今关于机器博弈和德州扑克算法的研究仍在蓬勃发展，不断开启新的篇章。

1.2 课题目的和意义

德州扑克是一项标志情商、智慧和勇气的牌类游戏，自 20 世纪初诞生便风靡全球，深受各个不同职业群体玩家的青睐。由于德州扑克非完备信息特性，游戏玩家都有未知的场上信息，同时还有诈唬等心理战术。这些特点使得德州扑克更贴近生活中真实的拍卖、谈判等金融政治博弈场景，所以德州扑克也许是测试

人工智能是否能够应付这些情景的一个平台。

前几年，AlphaGo 兴起了机器学习的热潮。在围棋这种完全信息的零和博弈中，作为算法需要解决的仅仅只是如何搜索大规模博弈树的问题，但是在德州扑克这种非完备信息博弈的问题中往往还藏有其他手段。而且对比围棋每次下棋有限的策略(19×19 个交叉点)德州扑克的策略几乎是无限的，专业选手往往押注在 1000 到十几万甚至百万美元，中间押注 10000\$ 和 10500\$ 也存在差距。所以由于德州扑克的种种特性，在 AlphaGo 4:1 战胜李世石的时候，同期的德州扑克还只能解决两人有限注的问题。

解决此类非完备信息博弈，如果解决的思路是从寻找纳什均衡出发，那么解决手段的基础就是 CFR(虚拟遗憾最小化算法)。此方法在 2007 年被提出后来被改进为 CFR+ 此后的德州扑克算法，比如前几年击败职业选手的冷扑大师，和今年的 Pluribus 都经过了 CFR 过程。

现代的博弈论快速与人工智能进行结合，形成了以数据驱动的博弈论新的框架。对于非完备信息博弈，其情况与完备信息博弈（如围棋）有很大的区别。博弈者不仅需要推理出他们对手到底知道什么，并且还要推理出他们什么样的行动会暴露自己的私人信息。解决非完备信息博弈的经典方法是计算或逼近所谓的纳什均衡。直到 2007 年，Martin Zinkevich 和其它阿尔伯塔大学计算机扑克研究小组的成员在 Michael Bowling 的领导下实现了一个突破，他们研究的虚拟遗憾最小化（Counterfactual regret minimization, CFR）便是这一次突破的产物。

扑克作为一类更为复杂和典型的非完备信息机器博弈问题。德州扑克是风靡全球的扑克游戏，许多专业的玩家通过游戏每年可赚得上百亿美元。德州扑克游

戏的信息不确定不完备、快速决策、重复博弈等特性集中了人工智能领域许多核心问题，这类游戏的研究对于整个人工智能领域有着极其关键的影响。研究实现高效的德州扑克算法无论在理论上还是现实应用中都具有重要的意义。而虚拟遗憾最小化算法通过从博弈的过程中吸取以往做决策的经验教训，多次迭代训练出高效的策略，以使智能体未来所做的动作最小遗憾。从该算法的直观含义上可以看出，这一算法模型的研究对解决现实生活中的金融股票、拍卖等经济问题十分有效，通过使每一轮的决策做到最小遗憾，尽可能的减少了自身的损失，从而使收益最大化。目前，世界上最有影响力的几大证券交易中心都将机器博弈部分纳入其交易系统，以此作为数据分析和决策支撑的重要工具。加拿大阿尔伯塔大学已将博弈中启发式搜索技术应用于土木工程任务，如在城市中找到设置下水道的最佳位置。游戏产业中也大量应用了机器博弈技术，如为大型网络游戏提供队友及智能 NPC，为棋牌类游戏提供代打机器人及陪练，为休闲游戏提供用户行为分析和建模，带来了巨大的商业价值。

自 2006 年起，世界年度扑克机器博弈大赛每年举办一次，并由人工智能国际协会（AAAI）举行，成立之初就成为数十个研究小组和业余爱好者的关注焦点。2015 年 Albert 大学 Bowling 等人对 CFR 进行改进，在两人有限注德州扑克取得的突破性的进展更是鼓舞了无数学者。因此研究 CFR 算法在非完备信息机器博弈中的应用具有广泛的理论研究价值和深远的实际应用意义。

二、国内外研究现状

2.1 国外研究现状

1959 年美国的塞缪尔(Samuel)设计了一个下棋程序，这个程序具有学习能力，它可以在不断的对奕中改善自己的棋艺。4 年后，这个程序战胜了设计者本人。又过了 3 年，这个程序战胜了美国一个保持 8 年之久的常胜不败的冠军。这个程序向人们展示了机器学习的能力，提出了许多令人深思的社会问题与哲学问题。随后机器博弈不断发展，并取得了显著成绩。1994 年计算机程序 Chinook 在跳棋上击败人类顶级玩家，成为第一个赢得人类冠军的机器算法；1997 年 IBM 开发的深蓝 (Deep Blue) 在国际象棋上击败 Kasparov，机器博弈领域发展如火如荼。在相继解决了西洋跳棋、奥赛罗等完备信息机器博弈之后，国外的研究者们逐渐将目光转向非完备信息博弈。非完备信息机器博弈自上世纪 90 年代开始成为人工智能领域研究的一个热点，特别是 1995 年 Billing 对扑克游戏的研究掀起了研究者们对扑克类机器博弈研究的热潮。

1997 年，Billing 等人成立了加拿大 Albert 大学计算机扑克研究组(University of Alberta Computer Poker Research Group, UACPRG)，并一度成为世界上最先进的扑克机器博弈问题的研究组织，产生了该领域的一大批学术成果，并发表了很多具有重要意义文章。这些文章大部分针对德州扑克的研究，并进一步开发出了对手建模和神经网络建模等技术。1998 年，第一个基于规则的德州扑克程序诞生，其智能水平已经达到人类玩家平均水平。2003 年，Darse 等人针对德州扑克游戏状态巨大的问题，首次提出了对游戏状态进行抽象的思路，根据此策略创建了具有较高水平的机器博弈程序，主要用于解决二人限注德州扑克问题，虽然

没有考虑对手信息，但已然具有极其重要的发展意义。2006 年 Billings 等人在二人限注德州扑克中将对手建模技术与博弈树搜索相结合，在合理的时间内可以实现深度优先搜索，再次实现了德州扑克的飞跃式发展。同年，世界年度扑克机器博弈大赛(Annual Computer Poker Competition, ACPC) 开始由 AAAI 和 IJCAI 会议联合举办，自 2006 年到 2014 年每年举办一次，从 2014 年开始改为每两年举办一次，并且成为非完备信息机器博弈领域的最权威评测机构。同时段，研究者们将纳什均衡逐渐引入德州扑克的研究。2007 年著名的扑克程序 Polaris 就是基于纳什均衡的。2007 年，Alberta 大学的 Martin Zinkevich 等人提出了基于遗憾最小化(Regret Minimization)的虚拟遗憾(Counterfactual Regret)的概念来解决大型博弈游戏。从前的遗憾最小化技术是通过计算平均整体遗憾值决定游戏中选择什么策略。传统上通常使用线性规划解决大型扩展式游戏，但显然，该方法不足以解决具有指数级状态数量的游戏。即使在小型的扩展式游戏中，如两人限注德州扑克，都有着多达 10^{17} 的游戏状态，因此在处理扑克游戏时线性规划方法需要结合抽样技术。如将游戏划分为一个个子游戏，独立实时的搜索子游戏的博弈树，或者通过 Hand-chosen 和 Automated 抽样方法可以使 10^{17} 的游戏状态减少到 10^7 ，并且同时保持较高的胜率。2007 年 Martin Zinkevich 等提出的 CFR 算法是将整体最小遗憾值分解到各个独立的信息集中计算局部最小遗憾值。实践证明，在大型的两人零和游戏中使用该方法近似纳什均衡，并且证明了庄家具有较大优势这一常识。2008 年，Polaris 在相关扑克游戏上首次击败了人类玩家，具有里程碑式的意义。2009 年 UACPRG 团队中的 Abou Risk 参加了当年的 ACPC 三人限注德州扑克比赛，并以显著性的优势赢得了当年三人有限注德州扑克比赛的冠军，

团队中的 Richard Gibson 自 09 年以来也一直致力于 CFR 算法的研究，他的团队利用该研究成果实现的智能体在 2010 年以来的连续 4 届 ACPC 大赛三人限注德州扑克项目中一直处于领先的地位。2015 年，来自 CMU 的 Sandholm 宣布最小规模的扑克游戏二人有限注德州扑克问题已经被解决。同刊中，Bowling 等人证明了利用其改进的 CFR+ 算法在二人有限注德州扑克问题上实现了弱拆解。尽管这个程序从每一局的概率上来说还是会输，但从长远来看可以取得统计学意义的胜利。

除使用 CFR 算法外，其他研究者也尝试过不同的方法求解德州扑克策略，均取得一定成果。2011 年荷兰马斯特里赫特大学的 Ponsen 率先将完备信息博弈中的蒙特卡洛树搜索方法应用在非完备信息的德州扑克上。2012 年葡萄牙波尔图大学的 Passos 等人将对手建模与强化学习算法相结合，为德州扑克研究提供了新思路。2014 年伦敦大学学院的 Heinrich 和 Silver[38] 将 Smooth UCT 算法应用在德州扑克上，结合了虚拟博弈和蒙特卡洛树搜索，在 Kuhn 扑克和 Leduc 扑克中收敛到近似纳什均衡。2016 年哥伦比亚大学和雅虎联合实验室的 Nikolai Yakovenko 等人使用深度学习基于人类玩家数据求解策略。

虽然很多研究机构在非完备信息机器博弈领域已经取得了丰硕的成果，同时也不断将其他学科的新鲜血液引入到了非完备信息机器博弈领域，但是还存在一定的不足。尤其在 CFR 算法的时间和存储开销方面一直是一个难以克服的难题。针对非完备信息机器博弈的研究技术还不够成熟，即使维奇已经被基本解决，关于德州扑克仍有很大研究提升空间。

2.2 国内研究现状

相较于机器博弈在国外的快速发展，国内的机器博弈起步较晚，近些年来也获得了越来越多的研究者的关注，也取得了一定的成就。国内在机器博弈方面更加关注完备信息分支，并取得了不错的成绩。东北大学是国内最早进行象棋机器博弈领域研究的院校之一，逐步整理出一套完整的棋盘表示、走法生成、基于自适应遗传算法的估值函数、开局库与残局库设计、参数优化等象棋博弈智能体体系。在 2006 年 8 月首届中国象棋人机大战中，东北大学开发的“棋天大圣”象棋程序与有中国象棋第一人称号的特级象棋大师许银川两战皆和。标志着象棋机器博弈已发展至较高的水平。

哈尔滨工业大学深圳研究生院计算机应用中心从 2003 年开始对非完备信息机器博弈进行研究，以四国军旗作为研究对象，在理论和实践上都取得了一定成果，某种程度上填补了国内非完备信息机器博弈研究的空白。2013 年哈工大课题组利用手牌评估和对手建模方法构建的德州扑克博弈程序，在 ACPC 大赛中的两人限制性德州扑克项目中荣获第四名；2014 年利用基于人工神经网络和风险占优策略搭建智能体决策模型在同年 ACPC 大赛 3 人 Kuhn 扑克项目中荣获第三名的佳绩。2016 年使用基于深度强化学习的方法在同年的 ACPC 的比赛中获得第八名。2017 年采用的基于抽象和游戏残局解算等算法开发的智能体在 2017 年 ACPC 大赛的二人非限制性德州扑克项目中荣获第三名。主要研究成果有非完备信息机器博弈系统的信息表示、启发函数优化算法、大规模博弈树搜索、对手建模、风险模型分析等。

三、课题要求

3.1 目的要求

1. 了解德州扑克的玩法和规则；
2. 理解 CFR 算法思路；
3. 熟悉 Python 语言编程；
4. 在两人有限注德州扑克中使用虚拟遗憾最小化（CFR）训练出强力 AI。

3.2 主要内容

1. 熟悉虚拟遗憾最小化（CFR）算法；
2. 熟悉德州扑克的一般思路；
3. 编写两人有限注德州扑克 AI。

3.3 预期目标

基于虚拟遗憾最小化实现两人有限注德州扑克算法。

四、课题研究进度安排

课题研究进度安排表

学期	周次	工作任务
2019-	1 周——3 周	文献翻译及开题报告
2020	3 周——6 周	完成工程主要框架，代码实现
第二学期	6 周——10 周	在服务器上测试并改进

五、主要参考文献

- [1] Science; Studies from University of Alberta Further Understanding of General Science (Heads-up limit hold'em poker is solved) [J]. Science Letter, 2015.
- [2] 代佳宁. 基于虚拟遗憾最小化算法的非完备信息机器博弈研究[D]. 哈尔滨工业大学, 2017.
- [3] 滕雯娟. 基于虚拟遗憾最小化算法的德州扑克机器博弈研究[D]. 哈尔滨工业大学, 2015.
- [4] 胡裕靖, 高阳, 安波. 不完美信息扩展式博弈中在线虚拟遗憾最小化[J]. 计算机研究与发展, 2014, 51(10): 2160-2170.

华中科技大学本科生毕业设计（论文）开题报告评审表

姓名		学号		指导教师	
院（系）专业					
<p>指导教师评语</p> <p>1. 学生前期表现情况。</p> <p>2. 是否具备开始设计（论文）条件？是否同意开始设计（论文）？</p> <p>3. 不足及建议。</p>					
<p>指导教师（签名）：</p> <p>年 月 日</p>					
教研室（系、所）或开题报告答辩小组审核意见					
<p>教研室（系、所）或开题报告答辩小组负责人（签名）：</p> <p>年 月 日</p>					