华中科技大学

本科生毕业设计（论文）参考文献译文本

Science; Studies from University of Alberta Further Understanding of General Science (Heads-up limit hold'em poker is solved)[J]. Science Letter,2015.

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

译文要求

一、 译文内容须与课题（或专业内容）联系，并需在封面注明详细出处。

二、 出处格式为

图书：作者.书名.版本（第×版）.译者.出版地：出版者，出版年.起页~止页

期刊：作者.文章名称.期刊名称，年号，卷号（期号）:起页~止页

三、 译文不少于5000汉字（或2万印刷符)。

四、 翻译内容用五号宋体字编辑，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。

五、 译文及其相应参考文献一起装订，顺序依次为封面、译文、文献。

六、 翻译应在第七学期完成。

译文评阅

导师评语

应根据学校“译文要求”，对学生译文翻译的准确性、翻译数量以及译文的文字表述情况等做具体的评价后，再评分。

评分：_____（百分制）　　指导教师（签名）：_____

年　　月　　日

两人有限注德州扑克

1. 前言

扑克是一类非完备信息的游戏，其中玩家对场上现有的游戏状态并不完全了解。尽管许多完备信息游戏已被解决（例如，“四子棋”和“跳棋”），但是很长时间内还没有人类玩家参与的非确定性非完备信息游戏被解决。而在本文中，两人有限注德州扑克博弈问题已基本解决。此外，本文的方法证明了游戏庄家具有较大优势这一基本常识。这一结果由改进的虚拟遗憾最小化（CFR+）实现，该算法能够解决比以前更大数量级的扩展式博弈游戏。

在计算机早期发展中，游戏便与博弈论和人工智能紧密相连。在计算方面，巴贝奇（Babbage）制定了一个非常详细，能够玩井字棋的“自动机”的计划，该计划可以用于分析机器下棋。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。半个多世纪以来，机器博弈一直是新算法思想的试验平台，在这方面取得的成就是人工智能发展的重要里程碑。比如，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，Watson 在 Jeopardy（一种博弈游戏）上击败 Jennings 和 Rutter。但是，击败顶尖的人类玩家并不等于“解决”游戏，而是计算出一种理论上可行的最优方案，该方案在所任何情况下都可以保持不败。彻底解决诸如“四子棋”和“跳棋”之类的游戏，实现了 AI 进步的里程碑。

迄今为止，人类解决的每一个游戏都是完备信息游戏。在完备信息游戏中，所有玩家在做出决策之前都会知晓当前场上的一切信息。象棋，西洋跳棋和西洋双陆棋等都是典型的完备信息游戏。在非完备信息的游戏中，玩家并不总是知道场上已发生的全部信息（例如，在桥牌游戏中玩家并不知道其他玩家的手牌，再比如拍卖时卖方并不能预测拍卖品价值）。这些游戏更具挑战性，其理论知识，计算方法和已解决的实例均落后于完备信息游戏。尽管大

部分室内游戏都是完备信息，但完备信息在现实世界的决策环境中却很少见。在与博罗诺夫斯基叙述的一次谈话中，现代博弈论的创始人冯·诺伊曼（von Neumann）得出了相同的看法:“现实生活中充满了虚张声势的欺骗和小战术，我不可能知道别人下一步的行动，如果想用某种理论解释日常生活的内在规律，那博弈论再恰当不过了。”

冯·诺依曼（Von Neumann）的看法与扑克游戏这一典型非完备信息游戏的策略不谋而合。 扑克游戏中每位玩家在游戏开始获得发给自己的手牌，然后轮流进行结构化下注，押注者可能拥有较大手牌（也可能是诈唬），然后跟对手下注或弃牌以放弃跟注。扑克在博弈论领域的早期发展中发挥了重要作用。 鲍罗（Borel）和冯·诺依曼（Von Neumann）的基础著作是通过发展扑克游戏中诈骗手段的数学原理而来的，小型扑克游戏在许多早期论文中很普遍。扑克也是世界上最受欢迎的纸牌游戏，全球有超过 1.5 亿玩家。

今天最受欢迎的扑克游戏是德州扑克。如果只有两名玩家，同时固定下注大小和加注次数（即有限注）进行比赛时，称为两人有限注德州扑克（HULHE，下文皆使用简称）。HULHE 被《教授，银行家和自杀之王》中记载的一系列高赌注游戏所普及。它也是人类竞争性扑克游戏的最小形式。HULHE 具有 3.16×10^{17} 个游戏可以达到的可能状态，大于四子棋而小于跳棋。 但是，由于 HULHE 是一种非完备信息游戏，因此在许多状态下玩家需要做出猜测，因为涉及到场上的未知信息（即对手的手牌）。 于是，游戏具有 3.19×10^{14} 个需要玩家自己决策的决策点。

尽管 HULHE 的信息量小于跳棋，但它的非完备特性使它成为对于计算机而言更具挑战性的游戏。Chinook 对抗世界跳棋冠军 Tinsley 赢得第一场比赛过去 17 年后，电脑程序北极星赢得了与职业扑克玩家的第一场比赛。尽管 Schaeffer 等人在 2007 年利用计算机解决了跳棋问题，然而两人有限注德州扑克问题仍未解决。这种缓慢的进展并非因为缺乏这方面的探索研究。扑克一直是人工智能，运筹学和心理学三大领域的难题，其研究历程可以追溯到 40

年前。17 年前，Koller 和 Pfeffer 就宣称：“我们离解决大型扑克之类的游戏还差得很远，而且我们目前也不太可能做到这一点。”以 HULHE 为例“大型扑克”的诞生始于 10 多年前，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。本文展示了为“全面解决”扑克游戏的持续研究成果。

Allis 对于游戏解决的程度给出了三个不同的定义。首先明确定义，我们称这种解决游戏的方式为“拆解”（solve）。对于二人（回合制、完全信息、不存在运气成分、有限）游戏，拆解分为以下几个层次：极弱拆解（Ultra-weakly Solved）：证明在初始局面下，双方的最大收益值是否可以确定存在 d，只需证明存在性即。弱拆解（Weakly Solved）：给出一个算法，能从初始局面开始保证先行方（或者后行方）取得最大收益，无论其对手采取何种策略。强拆解（Strongly Solved）：给出一个算法，能从任何局面开始给出双方的最强应对，即当前状态的任意一方都能获得最大收益。在非完备信息的游戏中，从一开始策略就不是不是唯一时，Allis 的“强拆解”方法就无法很好的实现。此外，由于玩家策略或游戏本身具有随机性，非完备信息游戏通常具有不同的实际收益值，而不是单一的结果（例如在国际象棋和跳棋中“胜局”，“败局”和“和局”）。最后，游戏的理论收益值通常是近似估计的，因此解决这类问题的一个关键因素是衡量解决方案的近似程度。弱拆解游戏近似估计的自然水平并不能得到精确的解，除非人类的在一局游戏中的时长足以建立一个基于统计学的方案。

我们在本文中已经取得了两人有限注德州扑克的弱拆解。此外，我们联系实际的收益值，证明了该游戏中庄家具有较大优势。

2. 解决非完备信息博弈游戏

扩展式博弈是典型的非完备信息博弈。这里的“博弈”泛指获益者和损失者之间的互动模型，在某些休闲游戏和工作场景十分常见，例如拍卖，谈判等。图 1 简单表述了 Kuhn 扑

克的起始步骤。扩展式博弈大都可以用树状图表示，它的分支表示游戏中可能存在的状态，即玩家采取的策略或执行后产生的结果。该树状图根据采取策略产生不同分支，每个分支都表示场上一名玩家采取的策略或当前游戏状态可能产生的结果。树状图的每个叶子节点表示特定的游戏过程直到结局，并且每个玩家都有各自的收益或亏损。与玩家相关的状态被划分为不同的信息集，若同一信息集包含多个游戏状态且这些游戏状态是无法分辨的，则是由于扩展式非完备信息博弈游戏中玩家无法获得确定的信息导致的[2]，扑克游戏中的信息集通常是玩家们无法区分的（例如发给对手的手牌）。信息集中各状态的分支是玩家的可采取行为，玩家为了最大化收益则针对每种信息状态建立模型，使得可采取策略符合某种概率分布。如果游戏中恰好有两个玩家，并且每个叶子节点处的玩家的收益值的和为零，则该游戏称为零和游戏。

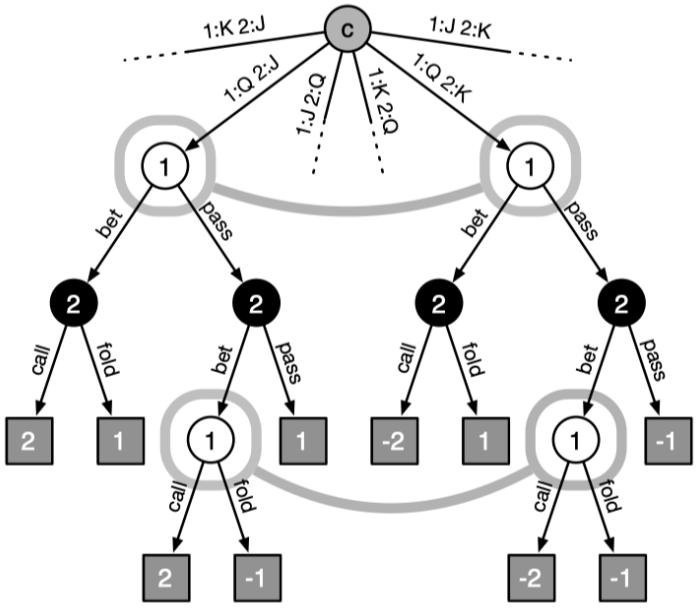

图 1：两人 Kuhn 扑克的树状图的一部分。玩家 1 被发牌 Q，而对手则被分配为牌 J 或牌 K。游戏状态指上图中，在每个信息集状态下，执行策略的玩家所标记的圆框（其中“c”表示信息集，并随机选择初始游戏状态）。箭头表示玩家可以选择的策略，并标注了游戏中的

含义。叶子节点是正方形，方框中是玩家 1 的最终游戏收益值（玩家 2 的收益值是玩家 1 取反）。 用粗灰线连接的状态是同一信息集的一部分，也就意味着，玩家 1 无法区分这一对状态，因为它们各自代表发给对手的另一张未观察到的卡牌。玩家 2 的状态也在信息集中，其中包含了此树状图未显示的状态。

博弈游戏的经典解决方法是达到纳什均衡，在两人博弈中，无论对方的策略选择如何，另一方都会选择某个确定的策略，则该策略被称作支配性策略。如果博弈双方的策略组合分别构成各自的支配性策略，那么这个组合就被定义为纳什均衡。达到纳什均衡时，每个博弈者的均衡策略都是为了达到自己期望收益的最大值，同时，其他玩家也遵循这样的原则，任何玩家都无法通过单方面违背该策略来增加其预期收益。所有有限的扩展式博弈都至少有一个纳什均衡，在零和博弈游戏中，玩家在所有均衡状态下的期望收益值都相同。一个ϵ-纳什均衡对每个玩家而言，是一种可以使得任何玩家都不能通过选择其他策略来增加自己的收益值是的策略。按照 Allis 的分类，如果能计算出零和博弈的收益值，那么这种情况属于极弱拆解；而如果能计算出纳什均衡策略，则是弱拆解。如果一个游戏计算出的纳什均衡值极小，甚至一个人终其一生玩过的无数次后都从统计上无法区分纳什均衡解的话，我们称该解为究极弱拆解。对于完备信息博弈游戏，求解纳什均衡通常涉及对游戏决策树的（部分）遍历。但是，相同的处理方法不能用于非完备信息博弈。我们简述了解决非完备信息博弈的发展现状，并通过在解决复杂度越来越大的扑克游戏方面所取得的进步来衡量算法，如图 2 所示。

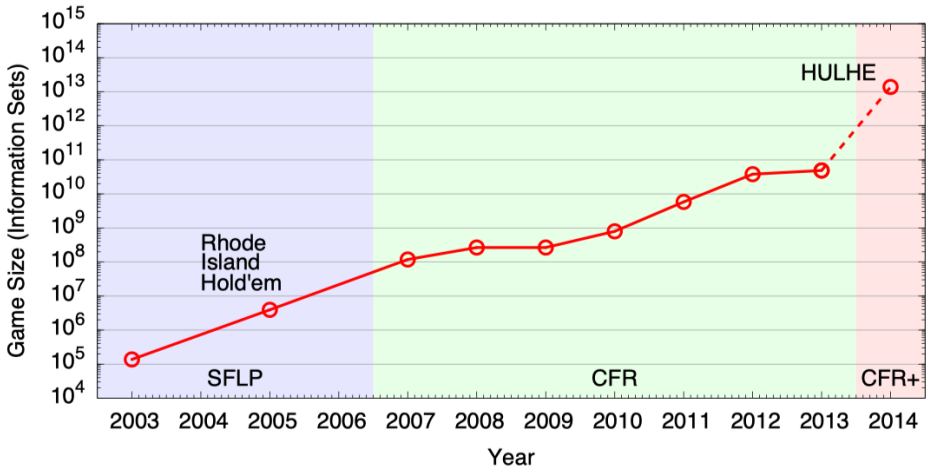

图 2：随着时间的推移，以独特信息集（即去除对称性后）衡量的非完备信息游戏的复杂度不断增加。不同颜色的区域代表所用的算法。虚线部分是本文获得的结果。

正则式线性规划：解决扩展式博弈游戏的最早方法是将其转换为正则博弈，将原始扩展式博弈中每对可能的策略值用矩阵表示，然后使用线性代数求解。然而，可确定的策略数量与游戏的信息集数量成指数关系。因此，尽管线性规划可以用数千种策略来处理正则式博弈，但即使只有几十个决策点也是不可行的。两人 Kuhn 扑克是一种有三张牌，一个下注回合和最大信息集数量为 12 的扑克游戏，这种情况可以用线性规划解决。 但是即使是 Leduc Hold'em，有 6 张牌，2 轮下注回合以及 2 次最大注（总共有 288 个信息集），也具有超过 10^{86} 种确定性策略。

序列式线性规划：Romanovskii 和 Koller 等人建立了解决非完备信息游戏的现代理论，提出一种用序列形式表示策略的方法。通过简单更改变量，他们得出扩展式博弈可以用线性形式表示并直接解决，而无需将指数形式转换为一般形式。序列式线性规划是解决扩展式非完备信息博弈的第一个算法，其计算时间随着表示游戏策略的多项式的增加而增加。2003 年，Billings 等人将 SFLP 首次应用于扑克，解决了 HULHE 一系列简化问题，并构建了第一个扑克竞赛的程序。2005 年，Gilpin 和 Sandholm 将 SFLP 与一种寻找游戏对称

性的自动化方法一起解决了 Rhode Island Hold'em，这是一款合成的扑克游戏，在去除对称性后有 3.94×10^6 个信息集。

虚拟遗憾最小化：2006 年，第一届年度计算机扑克大赛举办后，比赛中算法在解决复杂度较大的游戏方面不断取得进步，提出了多种改进方式。虚拟遗憾最小化（CFR）是目前最具有竞争性应用最广泛的算法。CFR 是一种通过两种遗憾最小化算法之间反复训练的过程来逼近扩展式博弈纳什均衡的迭代方法。通过多次迭代，将整体遗憾分解到各个独立的信息集中计算局部最小后悔值，后悔值是算法没有选择唯一的最佳策略而损失的游戏收益，只有在执行策略后才知道。遗憾最小化算法可以使后悔值随着游戏进行呈亚线性增长，最终实现与采用最优策略相同的效果。CFR 的关键点在于不会存储和最小化呈指数增长的决策数量的后悔值，而是存储和最小化在每个信息集和相应后续决策中的不断修改调整的后悔值，这样就可得出任何决策下的后悔值。通过在每一次迭代中对每个玩家的策略取平均值，可以近似求出纳什均衡，并且随着迭代次数的增加，近似程度也会提高。该算法所需的内存与信息集的数量是线性关系，而不是二次关系，高效的 LP 方法就应该是这样。由于之前解决大型博弈游戏通常会受到可用内存的限制，CFR 的出现使得可解决的游戏复杂度较之前大幅提升，比如 Koller 等人近期取得的成就。自 2007 年被提出后，CFR 已用于解决更为复杂的 HULHE 简化问题，在 2012 年可以处理具有 3.8×10^{10} 个信息集的博弈游戏。

3. 解决两人有限注德州扑克

两人有限注德州扑克完整的博弈树包含 3.19×10^{14} 个信息集。即使在消除游戏对称性之后，仍具有 3.19×10^{14} 个信息集（比以前解决的游戏大三个数量级），处理如此规模的博弈树结合 CFR 算法本身的特点，在实现过程中就面临两个挑战：存储和计算。在计算

过程中，CFR 算法必须存储每个信息集的最终解决方案和累加的后悔值。即使使用单精度（4 字节）浮点数，也需要 262 TB 的存储空间。此外，以往的经验表明，增加三个数量级的信息集，需要增加至少三个数量级以上的计算。为了解决这两个挑战，我们使用了本文共同作者最近提出的两个想法。

为了处理 CFR 算法在存储空间上面临的挑战，我们使用压缩的方法存储平均策略和累积的后悔值，同时使用定点计算的方法，设定定标因数，将其与所有的值首次相乘，结果截断为整数。生成整数进行排序，以达到压缩效率最大化，后悔值的压缩比约为 13:1，策略值的压缩比约为 28:1。这样，计算过程中只需要 11 TB 的存储空间来存储后悔值， 6 TB 的存储空间来存储平均策略值，分布遍及全部的计算节点。这样的规模存储在主存中是不切实际的，所以我们把值存储在每一个节点的本地磁盘中。每个节点负责一个子博弈集，是游戏决策树的一部分，子博弈树是根据公共可见的动作和自己的手牌以及底牌划分的，这样每个信息集都与一个子博弈相关。子博弈的后悔值和策略值使用流压缩技术从磁盘载入，更新，再存储回磁盘的，子博弈的部分需要的话，使用流媒体压缩技术进行解压缩和再压缩。通过使子博弈树变得足够大，更新时间主导子博弈过程的总时间。由于磁盘预缓存，会使效率下降的 5％。

为了解决计算方面的困难，来自 UACPRG 团队的 Michael Bowling 等人发明了一个叫做 CFR+的算法，这种算法是 CFR 的一个变体。CFR 算法每次迭代执行时仅仅抽样部分子博弈树进行更新，对每个信息集执行后悔匹配 regret matching 算法，为每个策略维持后悔值，选择后悔值为正数的策略，且动作的选择概率与后悔值成正比。相比之下，CFR+的迭代次数遍及整个博弈树，同时使用了后悔匹配 regret matching 算法的变形，即限定后悔值为非负数。后悔值为暂时负数的策略在证明有效后会被立即选择，而不是等多次迭代后其后悔值变成正数的时候才被选择。与 CFR 算法不同的是，策略值计算接近于 0 的期间，可

以凭借经验观察到玩家当前的策略利用情况，所以我们可以跳过计算和存储平均策略值得步骤，而是使用玩家当前的策略作为 CFR+的解决方案。根据以往的经验可以知道，即使计算平均策略值，这样计算方式也比蒙特卡洛 CFR 算法少了很多次计算，在计算过程中逐渐逼近纳什均衡，同时也非常适合大型博弈游戏的并行化处理。

与 CFR 算法一样，CFR+也是迭代算法，在计算过程中逐次逼近纳什均衡。近似程度通过其可用性衡量：与最坏情况下对手所采取的策略时相比，收益值小了多少。计算策略可用性需要计算最坏情况的收益值，通常需要遍历整个博弈树，因此长期以来，这对于 HULHE 这样复杂度很高的游戏来说都是不现实的。近期研究表明，可通过利用博弈游戏的非完备信息结构和收益规则来加速计算。这就是我们衡量纳什均衡策略值的近似程度的方法，并已在小型博弈游戏中进行了验证，而且针对 HULHE 中简单策略的可用性进行了独立计算。

我们通常根据预期值选择策略，但由于游戏具有随机性，即使全部考虑最坏情况，仍然不能保证在几次后可以获胜。对于一个游戏，如果适当的选择策略，从统计学上可以达到 95 %的置信度，我们就称其基本解决。假设一个人每天玩 12 小时，每小时玩 200 次扑克游戏，每天如此玩 70 年，然后该玩家考虑最坏情况下，对手所采取的收益最大化的策略并且自己零失误。玩家的总收入，作为数百万个单局游戏收益的总和，将服从正态分布。因此，在此扑克玩家的职业生涯中，收益总会有 1.64 的标准差或在 20 次游戏中至少有一次低于期望值（即策略值）。据报道，在 HULHE 中游戏收益的一般偏差约为 5 bb/g（每场比赛的大盲注，其中大盲注是 HULHE 的赌注单位），我们得出的阈值为

$(1.64 \times 5) / \sqrt{(200 \times 12 \times 365 \times 70)} \approx 0.00105$。因此，期望值在 1 M bb/g（每场游戏百万个大盲注）以下时的精确方案有很高的置信度，并且确实有 1/20 的机会可以在最坏情况下胜过对手。于是，1 M bb/g这一阈值宣布 HULHE 已基本解决。

4. 解决方案

我们的 CFR +算法是在 200 个计算节点的群集上实现的，每个计算节点具有 24 个 2.1 GHz AMD 内核，32GB RAM 和 1TB 本地磁盘。我们将游戏分为 110,565 个子游戏（根据底注和翻牌时的押注划分）。子游戏又分为 199 个工作节点，工作节点中的父节点负责博弈树的起始部分。工作节点采用并行方式更新，计算子节点的效用值然后回溯递归到父节点执行更新，平均花费 61 分钟才能完成一次迭代。该计算执行了 1579 次迭代，耗时 68.5 天，总共使用了 900 个核心计算量和 10.9 TB 的磁盘空间，其中包括大量文件的文件系统开销。

图 3 显示了随着计算量的增加，策略值的变化。该策略值可达到 0.986 M bb/g，意味着 HULHE 基本取得弱拆解。根据每个位置（庄家和非庄家）各自的策略值，我们可以得到游戏实际收益的界限：庄家的价值值在 87.7 和 89.7 M bb/g 之间，符合庄家在两人有限注德州扑克游戏中具备优势这一常识。

最终的策略会非常接近纳什均衡，也可以解决有关 HULHE 比赛中一些基本且长期存在的理论问题。图 4 展示了两个早期决策中的最终策略。人类玩家不同意用双手“ limp”（即呼唤是第一个动作而不是举手）。一般认为，该行为会迷惑对手使其放弃弃牌的机会，然后选择跟注。我们的解决方案证实了这一观点（请参见图 4a 中不存在蓝色）。该策略只拖延 0.06％的时间，而手牌下注时间不会超过 0.5％。在其他情况下，该策略展示了超出传统观点的一面，表明了人类可能会改善这一情况。该策略几乎不会在作为庄家的第一回合就梭哈（即赌上全部），而一些玩命的人类玩家则会以各种方式梭哈。即使握住最强的手牌（一对 A），该策略也会在 0.01％的时间进程内限制下注，最有可能加盖的手牌是一对 2，概率为 0.06％。也许更重要的是，该策略会选择（而不是弃牌）作为非庄家时比大多数人类玩家都会选择更激进的策略（参见图 4b 中相对较少的红色）。持有低排名的一对牌（例如三四）时，也有可能加注。

尽管以上这些观察仅是博弈论上最佳结果的一个示例（不同的纳什均衡可能产生不同效果），但它们已得到证实并且与人类当前对均衡博弈的认知相违背，说明人类可以从这种大规模博弈中学到很多关于游戏理论推理的知识。

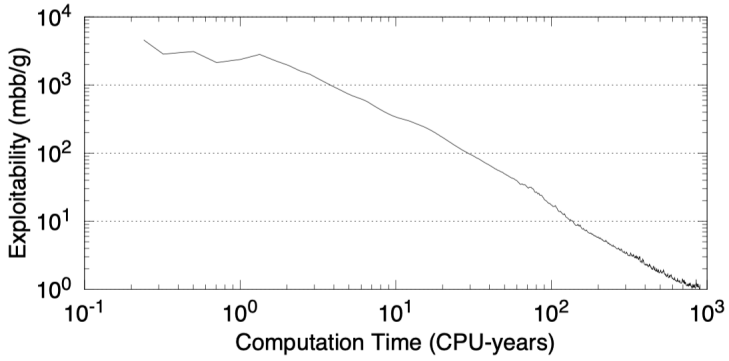

图 3：随着复杂度的增加，近似的策略值。 以每百万个大盲注点（M bb/g）衡量收益，是每次 CFR +迭代后采取当前策略收益值的衡量标准。 经过 1579 次迭代或 900 核心计算后，其策略值达到 0.986 M bb/g。

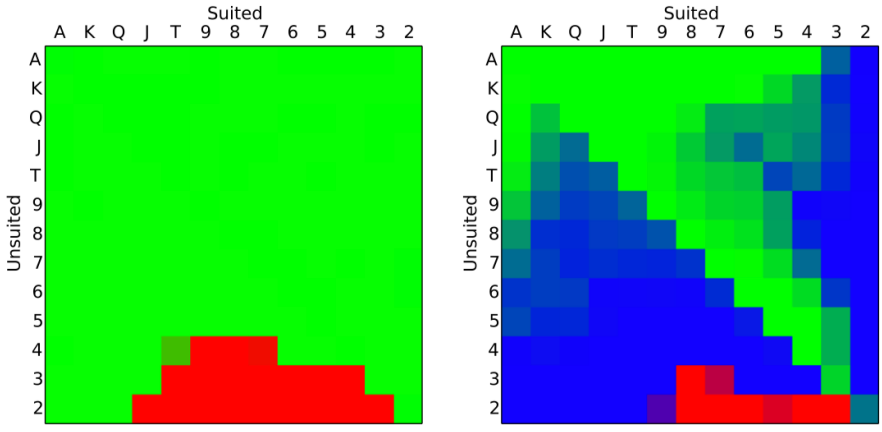

图 4：两个早期决策方案的策略概率。（A）发牌者在游戏中第一个动作的概率。（B）非庄家在庄家 limp 时的第一项行动的概率。 每个单元代表 169 张牌中的一对手牌（即玩家两张手牌），右上对角线由具有相同花色的牌组成，左下对角线由不同花色的牌组成。 单元格的颜色表示所采取的操作：红色表示弃牌，蓝色表示跟注，绿色表示加注，颜色的混合表示随机决策。

5. 总结

解决扑克游戏最重要的意义是什么？是为了实现算法上的的突破，这些突破可以使任何大规模博弈游戏的推理都变得更加容易。而且，尽管大部分时候博弈论看似只运用在棋牌，博弈和游戏方面，但在一些重要场合仍扮演着重要的角色[例如，它对冷战时期的政治格局的影响]。最近，涉及安全和隐私的的博弈论应用程序需求量越来越大，包括部署用于机场检查站，空中指挥调度中心和海上警卫队巡逻的系统。基于上述应用场景的 CFR 算法在没有意外干扰时已能够独立进行可靠的决策，并有可能进一步用于医疗卫生的决策。现实生活中的决策环境总会伴随着信息的不确定和丢失，因此算法的进步是投入到未来应用程序的关键。但是，我们也对图灵为自己的游戏作品的辩护作出回应:“我们不应该掩饰这样一件事实：完成一项工作的主要动机完全是寻找乐趣”。

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.

参考文献原文