1. 两人有限注德州扑克的CFR程序设计步骤

两人有限注德州扑克是一种典型的扩展式博弈，并且由于手牌，公共牌，可选取动作策略，以及诈唬等策略的存在，使得该游戏的信息集数量异常庞大，即使采取一些约束条件仍然很难降低其复杂度；设计CFR的思路就是基于扩展式博弈的博弈树，根据所选取的策略通过迭代构建新的节点，并且在每一个节点存储其包含的必要信息，到达叶子节点后就不断回溯至上一级节点，在遍历完一个节点的全部子节点后，更新当前节点的收益值和遗憾值以及历史策略等信息并存储，再回溯至上一级节点。本质上是一种深度优先遍历(Deep First Search, DFS)，完整的遍历完一次游戏的博弈树就完成了一次训练，在每次训练后更新全部信息集的遗憾值等信息，完成多次的训练后就得到一个具有一定智能的机器博弈体。以下是CFR的设计步骤：

1. 创建游戏以及两个玩家，各自获得两张手牌并且不知道对方的手牌，设置大小盲注，记录每个玩家下注的额度以及历史动作序列。每次函数的调用就相当于创建一个节点，选择一个当前的可执行动作加入历史动作序列，更新双方玩家的下注额，传入到下一个节点。
2. 判断当前回合数以及并执行是否更新公共牌，该处的执行函数只需要在迭代中判断一次，获得底牌即可，不用多次执行。
3. 判断当前节点是否为叶子节点，若是，则返回双方的收益值至上一级节点，否则执行第四步。
4. 判断当前回合是否结束，若是，则更新历史动作序列，进行下一轮的下注，否则继续当前回合。
5. 当前回合根据采取的策略不同建立新的节点并返回第二步，若所有子节点均遍历完，执行第六步
6. 遍历完当前节点的所有子节点后，将不同动作返回的收益值存储在当前节点。
7. 为防止被对手建模，根据轮盘赌转法选择策略，计算当前节点的平均收益值以及累积遗憾值。
8. 根据遗憾值匹配更新策略，即采取每一个动作的概率，并和本次遗憾值一块存储。
9. 将当前节点的两名玩家的收益值返回值上一级节点，执行第五步。
10. 两人有限注德州扑克游戏规则

德克萨斯扑克全称Texas Hold’em poker，中文简称德州扑克。它是一种玩家对玩家的公共牌类游戏。一张台面至少2人，最多10人，根据课题要求，本文以两人游戏为背景介绍德州扑克的规则。

德州扑克一共有52张牌，没有王牌。每个玩家分两张牌作为“底牌”，五张由荷官陆续朝上发出的公共牌。开始的时候，每个玩家会有两张面朝下的底牌。经过所有押注圈后，若仍不能分出胜负，游戏会进入“摊牌”阶段，也就是让所剩的玩家亮出各自的底牌以较高下，持大牌者获胜。游戏顺序是沿比赛桌，顺时针进行，庄家按钮的左手边开始。通常在按钮左手边的头两名玩家需要放下一个小盲注和一个大盲注，以此展开投注。两人德州扑克第一名玩家先下小盲注，称为玩家一（下同）；第二名玩家下大盲注，类似于庄家身份，称为玩家二，后手选取策略。

有限注区别于无限注，指的是玩家不能一次性将筹码全下（All-in），也就是赌博电影里的“梭哈”，因此不会在未到最后一轮下注前摊牌。有限注德州扑克保证玩家总是有足够的筹码进行押注，跟注，加注。相反，在非限注型玩法中，玩家可以选择全下，这意味着玩家已经将自己所有的筹码都投到底池中，没有剩余的筹码进行跟注、押注和加注了。选择 all-in 的玩家虽然不再参与押注，但是可以在摊牌阶段参与牌型的比较并决胜负，每局游戏结束后为所有玩家重置彩池。

游戏阶段包括：翻牌前（Pre-flop），翻牌（Flop，即头三张牌），转牌（Turn，第四张牌）和河牌（River,第五张牌）。

1. 翻牌前（Pre-Flop）：发底牌前，两位玩家分别下大、小盲注，荷官为各玩家发出两张底牌。从玩家一开始下注，该玩家有三种选择：弃牌（Fold）、跟注（Call）、加注（Raise）；然后玩家二开始下注，该玩家也有三个选择：弃牌（Fold），跟注（Call），加注（Raise）。跟注或加注都是依上一个玩家的下注额为参考。多人德州扑克中，发底牌后，由大盲注左侧的玩家开始依次押注行动。在没有人加注的情况下，大盲注位置的玩家行动结束后，本轮押注行动结束。若有人加注的情况下，直到最后加注玩家右则的第一个玩家行动结束后，本轮押注行动结束。
2. 翻牌（Flop）：荷官发出3张公共牌到牌桌中央，称为“翻牌”，所有人可见。第二轮下注从玩家一开始，此时有三种选择：弃牌（Fold）、看牌（Check）或下注（Bet）。后面的玩家可以选择弃牌、看牌（如果前面的玩家选择看牌）、跟注，加注、再加注。以后同第一轮下注。
3. 转牌（Turn）：荷官发出第4张公共牌到牌桌中央，称为“转牌”，所有人可见。本轮押注行动与翻牌一致。多了一张公共牌，各玩家的牌面可能发生转变，这时候各玩家需要重新判断牌局的形势。
4. 河牌（River）：荷官发出第5张公共牌到牌桌中央，称为“河牌”，所有人可见。本轮押注行动与翻牌一致。所有公共牌都发出了，这时候各玩家的牌面都固定下来，各玩家需再次判断形势作出合适的行动。
5. 四轮下注都完成后，若两名玩家中途均未弃牌，则进行比牌。比牌时，每位玩家用手中的两张底牌与五张公共牌中选出五张组成最大的牌型进行比较大小。比牌后，牌型最大的玩家赢得底池所有筹码。若最终牌型大小相同，各玩家只能平分底池。中途若有玩家弃牌，则牌局提前结束，奖池归另一名玩家。

五张牌牌型组合由大到小依次为：

1. 皇家同花顺（Royal Flush）：同一花色最大的顺子。
2. 同花顺（Straight Flush）：同一花色的顺子，牌面较大者胜，下同。
3. 四条（Four of a kind）：四张相同和一张单张。
4. 葫芦（Full House）：三张相同和一对。
5. 同花（Flush）：五张牌花色相同。
6. 顺子（Straight）：花色不同的顺子。
7. 三条（Three of a kind）：三张相同和两张单牌。
8. 两对（Two Pairs）：两对对子和一张单牌。
9. 一对（Pair）：一对对子和三张单牌。
10. 单张（Nothing）：五张单牌。
11. 正则博弈

正则（Norm form Game）形式是博弈论中描述博弈的一种方式。与扩展形式的博弈（Extensive Form Game）不同，正则形式不用图形来描述博弈，而是用矩阵来陈述博弈。与延展形式的表述方式相比，这种方式在识别出严格优势策略和纳什均衡上更有用，但会丢失某些信息。博弈的正则形式的表达中包括每个参与者所有显然的和可能的策略，以及和与其相对应的收益。

在完全信息的静态博弈（Static Games of Complete，Perfect Information）中，正则形式的表达形式是参与者的策略空间（Strategy Space）和收益函数（Payoff Function）。策略空间是某个参与者的所有可能策略的集合而策略是参与者在博弈的每个阶段——不管在博弈中这个阶段实际上是否会出现——将要采取的行动的完整计划。每个参与者的收益函数，是从参与者策略空间的向量积到该参与者收益集合（一般是实数集，数字表示基数效用或序数效用——在正则形式的表述方式中常常是基数效用）的映射。也就是说，参与者的收益函数把策略组合（所有参与者策略的清单）作为它的输入量，然后输出参与者的收益。

正则形式的博弈结构更准确的描述如下：假设有限的正则博弈游戏是一个包含如下要素的三元组:

1.有限集 表示玩家集合；

2.， 表示玩家可采取的所有动作的集合且是有限集，表示某一时刻游戏的策略集合即所有玩家在这一时刻采取某一动作的笛卡尔乘积，我们又把称为动作配置集；

3. 对于每个 , 表示效用函数: ，每个玩家都从其动作集合中选取一个动作，表示玩家在每个可能的动作配置集下所获得的收益。

如果并且，则游戏是零和的，否则游戏就是非零和的。这里强调的零和主要是指两人的游戏，对于三人游戏的 并不是零和。

如果，这样的游戏我们称为二人常和博弈，这样的游戏很容易转变成零和游戏且不需要改变游戏策略属性仅仅从某个玩家的收益中减去即可。

1. 扩展式博弈

与正则形式相对应，扩展形式是一种通过树来描述博弈的表达方式。每个节点（称作决策节点）表示博弈进行中的每一个可能的状态。博弈从唯一一个初始节点开始，通过由参与者决定的路径到达终端节点，此时博弈结束，参与者得到相应的收益。每个非终端节点只属于一个参与者；参与者在该节点选择其可能的行动，每个可能的行动通过边从该节点到达另一个节点。和正则形式不同，扩展形式允许互动的显式模型（Explicit Modeling of Interactions），互动中，一个参与者可以在博弈中多次行动，并且在不同的状态中可以做出不同的行为。当游戏玩家不是同时做决策而是顺序做决策时，或当游戏涉及非完备信息或随机事件，扩展式博弈比正则博弈更适用于为该类游戏建模。根据扩展式博弈的概念知道，几乎所有的棋牌类和扑克类的游戏都属于扩展式博弈游戏的范畴，因为这些游戏玩家执行动作均有先后顺序不是同一时间行动的。

扩展式博弈通常用树形结构表示，节点表示博弈状态也称决策节点，边表示玩家的策略选择，叶子节点代表游戏结束时每个玩家的收益。对于每个玩家来说，游戏状态被划分到信息集中，该玩家不可辨别的游戏状态被划分到同一信息集中。信息集是游戏状态的集合，扩展式博弈游戏形式化定义如下：

1. 有限集 表示参与博弈的玩家集合。
2. 序列的有限集表示已发生的历史动作序列，中的每个元素的前缀也位于中，每个不同的动作序列均代表一个决策节点。
3. 是终点状态集合，中每个元素均代表叶子节点。
4. 代表序列后可采取的动作集合。
5. 为每个非叶子节点后的可以行动的玩家。如果是，即确定行动玩家可以采取的策略。
6. 函数是在时当前节点关于的一种概率度量， 就是出现是动作出现的概率，其中，均为独立同分布。
7. 对于每个玩家，存在这样一个信息分区：，具有以下属性：若和 在均属于，则。建立一个信息集，对于任意，同时包含可采取动作以及下一名玩家等信息。本质上是同一名玩家具有相同后续可选取动作的信息集合，并存储可采取动作即下一名玩家等信息，用于在实际训练中存储更新策略。
8. 对于每个玩家，函数表示从叶子节点集合映射到实数集R上的收益值。定义为玩家的收益值范围。
9. 给定一个策略 为在信息节点处依据策略执行动作后到达节点的概率。
10. 纳什均衡

以最经典的囚徒困境为例，决策者可以利用下面的收益矩阵剔除劣势策略。囚徒困境的故事讲的是，两个嫌疑犯作案后被警察抓住，分别关在不同的屋子里接受审讯。警察知道两人有罪，但缺乏足够的证据。警察告诉每个人：如果两人都抵赖，各判刑一年；如果两人都坦白，各判八年；如果两人中一个坦白而另一个抵赖，坦白的放出去，抵赖的判十年。于是，每个囚徒都面临两种选择：坦白或抵赖。然而，不管同伙选择什么，每个囚徒的最优选择是坦白：如果同伙抵赖、自己坦白的话放出去，抵赖的话判一年，坦白比不坦白好；如果同伙坦白、自己坦白的话判八年，比起抵赖的判十年，坦白还是比抵赖的好。对竖排决策者来说，比较每列第一个数字可以发现，由于，因此无论横排决策者怎样选择，抵赖都是最优策略；同理，对于横排决策者来说，比较每列的第二个数字表明无论竖排决策者如何选择抵赖都能够获得更多的收益。因此，此博弈存在且唯一的纳什均衡是（抵赖，抵赖）。

|  |  |  |
| --- | --- | --- |
|  | 抵赖 | 坦白 |
| 抵赖 |  |  |
| 坦白 |  |  |

面对一群未知的对手时，如何计算游戏的策略配置文件是博弈者经常考虑的问题，对于这个问题最常见的解决方案就是纳什均衡。对纳什均衡更为直观的解释是：对于任意的，任意一名玩家单方面的违背策略，那么他获得的期望收益不高于 ，就说策略是纳什均衡的。

严格定义如下：对于任意的，策略是纳什均衡的，则

纳什均衡可以避免犯严格劣策略错误：不论对方采取什么策略，已方采取的这个策略所获得的收益都不能比其他策略得到更好收益。或者可以定义为在任何情况下，其他策略得到的收益都小于等于该策略。虽然在一两局的游戏中由于非完备信息游戏的随机特性纳什均衡策略未必最优，但从长远的统计学角度来看，纳什均衡策略一定是最优策略。

1. 虚拟遗憾最小化

在扩展式博弈中，遗憾匹配需要对所有的纯策略的进行存储，在上呈现指数分布，因此即使在中型的扩展式游戏中也是不可行的。因为在每一次迭代中更新策略文件的时间是呈指数分布的，同时对存储的要求也是呈指数增长的。每次迭代递归的访问信息集，根据遗憾匹配计算混合策略。在一般情况下，我们认为虚拟遗憾的计算就是指故意的使玩家的策略被限定在信息集 ，换一种说法就是，排除了真实的通过计算得来的玩家的数据的可能性。CFR 算法通过多次迭代，将整体遗憾分解到各个独立的信息集中计算局部最小遗憾值。CFR 算法的形式化如下：

代表博弈树中所有的叶子节点，，代表博弈树中的非叶子节点。表示玩家在叶子节点的效用值。定义在非叶子节点处虚拟价值：

玩家在节点不采取动作的虚拟遗憾值：

在信息集不采取动作的虚拟遗憾值：

玩家在次迭代中累加的虚拟遗憾值：

则对于信息集，通过后悔匹配得到当前的策略集合：

CFR 算法通过一次迭代在被访问到的每个信息集上计算虚拟价值和虚拟遗憾值，当下一次迭代开始时，将会带入上次迭代的结果通过遗憾匹配计算新的博弈策略。CFR 算法不断的迭代执行，递归的回溯博弈树进行计算并更新策略，从而最小化玩家的平均遗憾值。

CFR算法从随机策略开始，每次优化一个玩家的策略以提高其收益并反复迭代，最后取平均策略作为最终策略。每次优化用的是遗憾值最小化的办法，遗憾值就是事后最优选择的收益，减去当时选择的收益，遗憾值最小化就是把到目前为止的累计遗憾值拿过来，看哪一步累计遗憾值低，以后就多走这一步，为防止被对手建模，一般采用轮盘赌转法随机选择策略。对于两人零和游戏，可以证明改进的CFR+会收敛到纳什均衡点。

1. 实验验证

采用Python语言以及PyCharm编辑器。由于Python的面向对象特性因此特别适合编写博弈游戏的流程。在确定整体框架后就开始编写。首先确定类，游戏平台，玩家，牌桌，扑克牌，荷官，动作，策略，牌级（摊牌后手牌与公共牌的最大牌力组合），牌力（两张手牌）等。

在未加入CFR策略时用玩家-玩家对战测试程序，无误后再引入CFR算法计算游戏策略并训练高智能体。由于是两人对局，因此对局过程相对简单。首先创建一局游戏，玩家个数默认为2，给每人分配一定的筹码，并分别下大小盲注；创建荷官和牌桌并生成52张扑克牌，随机洗牌后，荷官给两名玩家发两张手牌并出示三张公共牌，根据玩家不同的对局状态（翻牌前，翻牌，转牌，河牌）预先设置好可选择的动作集合（看牌，下注，跟注，加注，弃牌），玩家在不同阶段选择。如果有一人选择弃牌，则另一名玩家获胜。如果没有人弃牌，则河牌后比较两人牌力大小，牌力大者获胜。