



(12) 发明专利申请

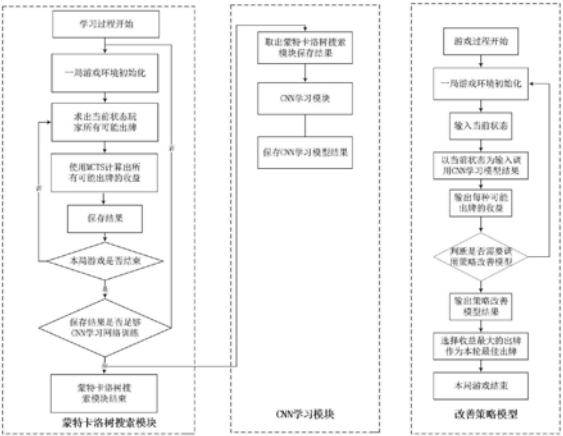
(10) 申请公布号 CN 111729300 A
(43) 申请公布日 2020. 10. 02

(21) 申请号 202010589925.X
(22) 申请日 2020.06.24
(71) 申请人 贵州大学
地址 550025 贵州省贵阳市花溪区花溪大道
(72) 发明人 王以松 彭啟文
(74) 专利代理机构 北京众合诚成知识产权代理有限公司 11246
代理人 刘妮
(51) Int. Cl.
A63F 13/46 (2014.01)
A63F 13/822 (2014.01)
G06N 3/04 (2006.01)
G06N 3/08 (2006.01)

权利要求书2页 说明书7页 附图4页

(54) 发明名称
基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法
(57) 摘要

本发明公开了机器学习技术领域的基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法,所述的方法包括如下步骤:随机开始游戏并在每个玩家出牌时,以玩家当前状态为根节点,玩家按照斗地主规则可能采用的动作作为根节点的直接子节点;再从博弈树的根节点开始,使用蒙特卡洛树搜索算法进行不断的模拟抽样学习;当使用蒙特卡洛树搜索算法获得的数据足够多时,以(状态及可能出牌,当前状态下可能出牌对应的收益)为数据样本,不断训练卷积神经网络CNN学习网络,直到该网络稳定为止;针对CNN网络在学习时可能存在的误差,进一步使用策略改善算法对CNN网络学习结果进行修正改善。



1. 基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法, 其特征在于: 所述的方法包括如下步骤:

随机开始游戏并在每个玩家出牌时, 以玩家当前状态为根节点, 玩家按照斗地主规则可能采用的动作作为根节点的直接子节点;

再从博弈树的根节点开始, 使用蒙特卡洛树搜索算法进行不断的模拟抽样学习;

当使用蒙特卡洛树搜索算法获得的数据足够多时, 以 (状态及可能出牌, 当前状态下可能出牌对应的收益) 为数据样本, 不断训练卷积神经网络CNN学习网络, 直到该网络稳定为止;

针对CNN网络在学习时可能存在的误差, 进一步使用策略改善算法对CNN网络学习结果进行修正改善。

2. 根据权利要求1所述的基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法, 其特征在于: 所述状态包括但不限于每个玩家已出牌, 玩家手牌张数, 当前玩家手牌。

3. 根据权利要求1所述的基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法, 其特征在于: 所述蒙特卡洛树搜索算法为蒙特卡洛方法和博弈树搜索相结合的方法。

4. 根据权利要求3所述的基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法, 其特征在于: 所述蒙特卡洛方法是利用经验平均来代替随机变量的期望, 具体方法为:

通过蒙特卡洛方法获取到一系列收益 $G_1(s), G_2(s), \dots, G_n(s)$ 。根据大数定律, 当 n 趋于无穷大时, 抽样收益的均值趋近于期望值, 定义 $v(s)$ 为系列收益的平均值, 即:

$$v(s) = \frac{G_1(s) + G_2(s) + \dots + G_n(s)}{n} \quad (1)$$

$$v(s) \rightarrow v_{\pi}(s) \text{ as } n \rightarrow \infty \quad (2)$$

其中, s 为游戏状态, $v_{\pi}(s)$ 为游戏 s 状态时的期望值。

5. 根据权利要求3所述的基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法, 其特征在于: 所述博弈树搜索算法包括以下步骤:

扩展节点的选择: 递归地应用节点选择函数, 从所有待选择的节点中, 选择一个节点作为本次扩展的根节点, 从该节点开始, 对该节点表示的博弈局面进行一次模拟;

扩展步骤: 将一个或多个节点增加到MCTS搜索树中, 普通的策略是每次迭代, 只向博弈树种增加一个新节点;

模拟: 模拟实际玩家博弈过程, 进行从本次开始模拟的节点到终止状态的一次博弈过程;

反馈: 模拟的结果会从本次模拟的终止节点开始, 通过逐层反馈给父节点的方式, 最终将模拟结果返回给根节点。

6. 根据权利要求1所述的基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法, 其特征在于: 所述扩展节点的选择使用的算法为UCT算法, 所述UCT算法为:

$$\gamma_i = \bar{V}_i + C \sqrt{\frac{2 \ln \sum_j n_j}{n_i}} \quad (3)$$

式中 γ_i 表示节点 i 的选择评估值, \bar{V}_i 表示节点 i 的平均收益; C 是常数, 其作用是为了平衡探索和利用; n_i 是以第 i 个节点作为模拟搜索的根节点的次数。

7. 根据权利要求1所述的基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法, 其特征在于: 所述蒙特卡洛树搜索算法在满足最大抽样次数或者达到时间耗尽等设置后, 根据博弈树中第一层子节点中每个节点的估值, 从中选择一个决策作为本次MCTS算法的最佳决策。

8. 根据权利要求1所述的基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法, 其特征在于: 所述CNN学习网络由4层卷积层和3层全连接层构成, 在其中还增加了3层池化层并使用Relu作为激活函数, 该网络使用当前状态及某一种可能的出牌作为输入, 其输入大小为 $15 \times 15 \times 29$, 输出为当前状态及某一种可能出牌的收益。

9. 根据权利要求1所述的基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法, 其特征在于: 所述策略改善方法实质是蒙特卡洛树搜索算法, 通过给CNN学习网络输入某一状态及该状态下的一种可能出牌后, CNN学习网络会输出当前状态下该可能出牌的收益值, 通过循环每种可能出牌可求解出当前状态下每种可能出牌的收益; 然后将当前状态下每种可能出牌的收益值作为策略改善模块中蒙特卡洛树搜索的初始值, 在一定时间范围内, 再对当前状态进行不断抽样模拟, 修正CNN学习网络中可能存在的误差。

基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法

技术领域

[0001] 本发明涉及机器学习技术领域,具体为基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法。

背景技术

[0002] 近年来,随着机器学习的发展,该方法也在完备信息博弈方面取得了显著的成果。其中具有里程碑意义的是:2016年3月15日,谷歌公司使用深度学习和强化学习等方法,开发的AlphaGo在围棋领域打败了世界围棋冠军李世石,其标志机器在围棋领域实现了超人的表现。随后AlphaGo Zero的训练完全是通过自我学习来进行,AlphaGo Zero从随机游戏开始,无需任何监督或使用人类数据,并且仅使用棋盘上的黑白子作为原始输入特征,通过不断地学习最后以显著优势胜过使用人类数据参与训练的AlphaGo。与机器学习在围棋、象棋等棋盘游戏中取得的傲人成果相比,在诸如德州扑克、“斗地主”等纸牌游戏并未实现较好的表现。

[0003] 研究者对于纸牌游戏的研究,主要集中在德州扑克这种不完备信息博弈游戏中,并取得了不错的成果,如卡耐基梅隆大学开发的Libratus和阿尔伯塔大学开发的DeepStack都能在与人类玩家进行一对一的模式中取得不错的结果。其中Libratus假设两玩家在每次决策都使用Monte Carlo Counterfactual Regret Minimization(MCCFR)求解得到,以使得所求策略趋近于纳什均衡,但是纳什均衡等博弈知识大部分都是建立在非合作博弈之上,目前对于像“斗地主”游戏中农民间体现出的合作博弈暂无较好的解决方案;而DeepStack则使用了深度学习的方法。此外,不论是Libratus或DeepStack,它们都只能进行一对一游戏。

[0004] 相较于德州扑克的研究,在“斗地主”游戏方面的研究较少,其中上海交通大学的YangYou等人提出一种Combinational Q-Learning(CQL)的解决方案,该方案在于原始深度强化学习如DQN、A3C的比较中,取得了显著效果(其中原文实验表明:在“斗地主”游戏中,DQN算法并未收敛)。但是在与人类进行实际游戏的过程中,效果往往没有那么好。

[0005] 斗地主,作为在中国比较流行的纸牌游戏,深受大众的喜欢。在2018年腾讯公司年度“斗地主”锦标赛中,参与人数多达8000万。与之相反的是,现在对于“斗地主”的研究较少,主要因为其较难且对其重视度不够。斗地主是一款3人的纸牌游戏,对“斗地主”游戏规则属于现有技术。其中研究难主要体现在如下方面:(1)在游戏过程中,扑克信息是部分隐藏的;(2)按照“斗地主”游戏规则,游戏过程中游戏空间较大;(3)该游戏是多人博弈的问题,在游戏中还体现了合作博弈。

[0006] 基于此,本发明设计了基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法,以解决上述提到的问题。

发明内容

[0007] 本发明的目的在于提供基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方

法,以解决上述背景技术中提出的问题。

[0008] 为实现上述目的,本发明提供如下技术方案:基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法,所述的方法包括如下步骤:

[0009] 随机开始游戏并在每个玩家出牌时,以玩家当前状态为根节点,玩家按照斗地主规则可能采用的动作作为根节点的直接子节点;

[0010] 再从博弈树的根节点开始,使用蒙特卡洛树搜索算法进行不断的模拟抽样学习;

[0011] 当使用蒙特卡洛树搜索算法获得的数据足够多时,以(状态及可能出牌,当前状态下可能出牌对应的收益)为数据样本,不断训练卷积神经网络CNN学习网络,直到该网络稳定为止;

[0012] 针对CNN网络在学习时可能存在的误差,进一步使用策略改善算法对CNN网络学习结果进行修正改善。

[0013] 优选的,所述状态包括但不限于每个玩家已出牌,玩家手牌张数,当前玩家手牌。

[0014] 优选的,所述蒙特卡洛树搜索算法为蒙特卡洛方法和博弈树搜索相结合的方法。

[0015] 优选的,所述蒙特卡洛方法是利用经验平均来代替随机变量的期望,具体方法为:

[0016] 通过蒙特卡洛方法获取到一系列收益 $G_1(s), G_2(s), \dots, G_n(s)$ 。根据大数定律,当 n 趋于无穷大时,抽样收益的均值趋近于期望值,定义 $v(s)$ 为系列收益的平均值,即:

$$[0017] \quad v(s) = \frac{G_1(s) + G_2(s) + \dots + G_n(s)}{n} \quad (1)$$

$$[0018] \quad v(s) \rightarrow v_\pi(s) \text{ as } n \rightarrow \infty \quad (2)$$

[0019] 其中, s 为游戏状态, $v_\pi(s)$ 为游戏 s 状态时的期望值。

[0020] 优选的,所述博弈树搜索算法包括以下步骤:

[0021] 扩展节点的选择:递归地应用节点选择函数,从所有待选择的节点中,选择一个节点作为本次扩展的根节点,从该节点开始,对该节点表示的博弈局面进行一次模拟;

[0022] 扩展步骤:将一个或多个节点增加到MCTS搜索树中,普通的策略是每次迭代,只向博弈树种增加一个新节点;

[0023] 模拟:模拟实际玩家博弈过程,进行从本次开始模拟的节点到终止状态的一次博弈过程;

[0024] 反馈:模拟的结果会从本次模拟的终止节点开始,通过逐层反馈给父节点的方式,最终将模拟结果返回给根节点。

[0025] 优选的,所述扩展节点的选择使用的算法为UCT算法,所述UCT算法为:

$$[0026] \quad \gamma_i = \bar{V}_i + C \sqrt{\frac{2 \ln \sum_j^n n_j}{n_i}} \quad (3)$$

[0027] 式中 γ_i 表示节点 i 的选择评估值, \bar{V}_i 表示节点 i 的平均收益; C 是常数,其作用是为了平衡探索和利用; n_i 是以第 i 个节点作为模拟搜索的根节点的次数。

[0028] 优选的,所述蒙特卡洛树搜索算法在满足最大抽样次数或者达到时间耗尽等设置后,根据博弈树中第一层子节点中每个节点的估值,从中选择一个决策作为本次MCTS算法

的最佳决策。

[0029] 优选的,所述CNN学习网络由4层卷积层和3层全连接层构成,在其中还增加了3层池化层并使用Relu作为激活函数,该网络使用当前状态及某一种可能的出牌作为输入,其输入大小为 $15*15*29$,输出为当前状态及某一种可能出牌的收益。

[0030] 优选的,所述策略改善方法实质是蒙特卡洛树搜索算法,通过给CNN学习网络输入某一状态及该状态下的一种可能出牌后,CNN学习网络会输出当前状态下该可能出牌的收益值,通过循环每种可能出牌可求解出当前状态下每种可能出牌的收益;然后将当前状态下每种可能出牌的收益值作为策略改善模块中蒙特卡洛树搜索的初始值,在一定时间范围内,再对当前状态进行不断抽样模拟,修正CNN学习网络中可能存在的误差。

[0031] 与现有技术相比,本发明的有益效果是:

[0032] 针对行为空间大和玩家间关系导致的问题,本发明使用蒙特卡洛树搜索方法,对行为空间进行抽样模拟以期通过不断模拟后,能求解出每种可能出牌动作的收益值,并从中选择出最大收益值对应的动作作为本轮的最佳出牌动作。在本发明中,通过使用MCTS进行不断游戏,并将其结果作为原始数据,提供给卷积神经网络进行学习,以期通过给CNN网络提供游戏状态,计算出当前游戏状态下每种可能出牌的收益,再从中选择收益最大的出牌作为本轮的最佳出牌动作,从而大幅缩短决策时间。

附图说明

[0033] 为了更清楚地说明本发明实施例的技术方案,下面将对实施例描述所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0034] 图1为本发明方法流程图;

[0035] 图2为本发明MCTS博弈树搜索算法过程示意图;

[0036] 图3为本发明CNN网络学习框架图;

[0037] 图4为本发明L1损失函数得到的损失变化图;

[0038] 图5为本发明作为地主时算法和RHCP算法胜率变化图;

[0039] 图6为本发明作为农民时算法和RHCP算法胜率变化图。

具体实施方式

[0040] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0041] 请参阅图1,本发明提供一种技术方案:基于蒙特卡洛树搜索和卷积神经网络斗地主策略研究方法,包括如下步骤:

[0042] 随机开始游戏并在每个玩家出牌(决策)时,以玩家当前状态(每个玩家已出牌,玩家手牌张数,当前玩家手牌等)为根节点,玩家按照斗地主规则可能采用的动作作为根节点的直接子节点;

[0043] 再从博弈树的根节点开始,使用MCTS算法进行不断的模拟抽样学习;

[0044] 计算出当前状态下,每种可能出牌的收益值并将其保存,不断重复该过程;

[0045] 当使用MCTS算法获得的数据足够多时,以(状态及可能出牌,当前状态下可能出牌对应的收益)为数据样本,不断训练CNN学习网络,直到该网络稳定为止;

[0046] 针对CNN网络在学习时可能存在的误差,进一步使用策略改善算法对CNN网络学习结果进行修正改善。

[0047] 其中,针对问题规模大的问题,相关研究人员将蒙特卡洛方法和博弈树搜索相结合提出了MCTS算法,具体为蒙特卡洛树搜索(Monte Carlo tree search,MCTS)方法,该方法保证在降低问题规模的同时,能保持所求解的近似最优性。

[0048] 蒙特卡洛方法是利用经验平均来代替随机变量的期望。比如在游戏中的状态s时期望值为 $v_{\pi}(s)$,该期望值难以通过计算求出,但是可以通过蒙特卡洛方法获取到一系列收益 $G_1(s), G_2(s), \dots, G_n(s)$ 。根据大数定律,当n趋于无穷大时,抽样收益的均值趋近于期望值。定义 $v(s)$ 为系列收益的平均值,即:

$$[0049] \quad v(s) = \frac{G_1(s) + G_2(s) + \dots + G_n(s)}{n} \quad (1)$$

$$[0050] \quad v(s) \rightarrow v_{\pi}(s) \text{ as } n \rightarrow \infty \quad (2)$$

[0051] MCTS算法以初始化搜索树开始,一般抽象当前的游戏状态作为一个单独的博弈树根节点,一旦博弈树初始化完成后,就会在规定的时间内重复如图2所示的MCTS博弈树搜索算法过程,搜索过程可以分为4个步骤:

[0052] 扩展节点的选择:递归地应用节点选择函数,从所有待选择的节点中,选择一个节点作为本次扩展的根节点,从该节点开始,对该节点表示的博弈局面进行一次模拟。

[0053] 扩展步骤:将一个或多个节点增加到MCTS搜索树中。普通的策略是每次迭代,只向博弈树种增加一个新节点。

[0054] 模拟:模拟实际玩家博弈过程,进行从本次开始模拟的节点到终止状态的一次博弈过程;

[0055] 反馈:模拟的结果会从本次模拟的终止节点开始,通过逐层反馈给父节点的方式,最终将模拟结果返回给根节点。

[0056] MCTS算法会在满足最大抽样次数或者达到时间耗尽等设置后,根据博弈树中第一层子节点中每个节点的估值,从中选择一个决策作为本次MCTS算法的最佳决策。

[0057] 在扩展节点选择的时候,所使用的算法存在很大差异。其中,现在广泛使用的是2006年,Kocsis和Szepesvari在UCB算法基础上,针对MCTS算法中扩展节点选择问题,提出的UCT算法。UCT算法为:

$$[0058] \quad \gamma_i = \bar{V}_i + C \sqrt{\frac{2 \ln \sum_j^n n_j}{n_i}} \quad (3)$$

[0059] 在公式3中 γ_i 表示节点i的选择评估值, \bar{V}_i 表示节点i的平均收益;C是常数,其作用是为了平衡探索和利用; n_i 是以第i个节点作为模拟搜索的根节点的次数。实践证明该算法非常高效。

[0060] 其中,使用蒙特卡洛树搜索模块作为策略进行博弈时,虽然效果不错,但是每局游戏开始的时候,都需要将当前状态作为根节点进行一次MCTS,并且在使用MCTS进行模拟采样的过程中,为了保证准确率,往往每一轮游戏也需要耗费几分钟时间,这在实际游戏过程中显然时无法接受。其外,每轮游戏中,使用MCTS花费大量时间计算出结果却没有被有效得使用,针对这种情况,本发明使用卷积神经网络(CNN)学习MCTS搜索后的结果,充分利用MCTS模块的结果,达到缩短博弈决策时间的目的。

[0061] CNN网络由4层卷积层和3层全连接层构成,此外在其中还增加了3层池化层并使用Relu作为激活函数。该网络使用当前状态及某一种可能的出牌作为输入,其输入大小为15*15*29,输出为当前状态及某一种可能出牌的收益。其中输入具体表示如下:

[0062] X维度:X维度中的每个位置,分别顺序表示{3,4,5,6,7,8,9,T,J,Q,K,A,2,L,B}等15种扑克。

[0063] Y维度:Y维度中的(下标)表示对应X维度中的扑克张数,其中(0-4张分别用0000、0001,0010,0100,1000表示),Y维度中的4-13(下标)分别表示出牌类型single、pair、three、three_plus_one、three_plus_two、sequence_one、sequence_two、sequence_three、bomb、k_bomb等10种,14(下标)表示该出牌是否为地主玩家。

[0064] Z维度如下表:

[0065]	维度	意义
	0-2	分别表示当前玩家、下家、上家8轮之前所有出牌的张数,不区分每次出牌类型。
	3-5	分别表示当前玩家、下家、上家前第8轮的出牌
	6-8	分别表示当前玩家、下家、上家前第7轮的出牌
[0066]	9-11	分别表示当前玩家、下家、上家前第6轮的出牌
	12-14	分别表示当前玩家、下家、上家前第5轮的出牌
	15-17	分别表示当前玩家、下家、上家前第4轮的出牌
	18-20	分别表示当前玩家、下家、上家前第3轮的出牌
	21-23	分别表示当前玩家、下家、上家前第2轮的出牌
	24-26	分别表示当前玩家、下家、上家前第1轮的出牌
	27	表示本轮玩家出牌
	28	表示当前玩家手牌,只记录牌张数,不区分牌型

[0067] 如图3所示,为CNN网络学习框架,其中包含4个卷积层、3个池化层以及3个全连接。

[0068] 使用CNN网络学习模型对于MCTS所保存的500多万条记录进行学习,使用L1损失函

数得到其损失变化如下图4所示图中x轴表示epoch,y轴表示收益损失。从图可以看出,在0~30epoch时,收益损失出现大幅度下降,在30epoch之后,收益损失变化不大,最终在0.033附近浮动。

[0069] 其中,经过蒙特卡洛树搜索模块学习到的策略,并使用CNN网络对该策略进行学习后,CNN学习网络可以直接进行“斗地主”游戏。但是使用MCTS模块进行学习的时候,由于时间限制从而导致有些决策分支未能进行充分的探索,加之其后使用CNN学习网络学习策略时,不可避免的存在误差。鉴于此,引入策略改善方法。

[0070] 策略改善方法,其实质就是蒙特卡洛树搜索。在博弈的过程中,通过给CNN学习网络输入某一状态及该状态下的一种可能出牌后,学习网络会输出当前状态下该可能出牌的收益值,通过循环每种可能出牌可求解出当前状态下每种可能出牌的收益;然后再以此结果作为策略改善模块中蒙特卡洛树搜索的初始值,在可能的时间范围内,再对当前状态进行不断抽样模拟,以修正CNN学习网络中可能存在的误差。此外,CNN学习网络本质是学习MCTS模块中学习到的策略,因此在CNN网络的基础上再次使用MCTS进行抽样模拟,实质就是在MCTS模块的抽样结果基础上,在限制时间范围内增加抽样次数,使得抽样的结果更趋近真实结果。

[0071] 实验结果:

[0072] 首先简单介绍目前可供测试的“斗地主”智能体,其后分别与本发明实现的“斗地主”智能体做直接或间接的比较。

[0073] RCHP算法:其总体思路是将手牌按照“斗地主”规则进行不同的组合,并选择使得出牌后手牌价值较高的出牌作为本轮最佳出牌。

[0074] Naive DQN和A3C:借鉴深度强化学习在游戏中的经验,使用其思想应用到“斗地主”游戏中。

[0075] CQL:该算法是上海交通大学YangYou等人所提出。

[0076] 与经典的深度强化学习算法比较:

[0077] 游戏设定:使用RCHP作为农民玩家的决策算法,分别于其它算法作为农民进行博弈,如下表。

[0078]	算法	胜率
	NaiveDQN	15%左右
	A3C	接近20%
	CQL	接近50%
	本发明算法	65.6%

[0079] 表中Naive DQN,A3C的数据是直接引用上海交通大学YangYou等人论文中结果。

[0080] 从表中可以看出,本发明算法作为地主与RHCP算法博弈时,效果相较于经典深度强化学习有较大优势。

[0081] 与RHCP算法比较:

[0082] 游戏设定:在进行博弈的过程中,通过程序随机的将51张扑克平均的分配给每个玩家(发牌过程,每玩家分配到17张手牌),然后随机指定一个玩家作为地主,再将剩余的3张地主牌亮出并分配给该地主玩家。从地主玩家开始,按照“斗地主”博弈规则进行博弈。在博弈过程中:一种角色的决策算法为MCM算法、另一种角色的决策算法为RHCP算法;其中两

农民的决策算法均为农民角色的决策算法。根据上述游戏设定,本发明算法作为地主和农民分别与RHCP算法进行500局游戏,结果如下:

[0083] 1、当本发明算法作为地主时,本发明算法和RHCP算法相比,胜率变化图如图5所示,从图中可以看出,本发明算法作为地主时,胜率为65.6%,明显高于RHCP算法的约34.4%。

[0084] 2、当本发明算法作为农民时,本发明算法和RHCP算法相比,胜率变化图如图6所示,从图中可以看出,本发明算法和RHCP算法作为农民时,总胜率大约为57%,而RHCP算法作为地主,取得的胜率为43%左右。

[0085] 3、不区分角色时,本发明算法胜率为61.3%,而RHCP算法的胜率为38.7%。

[0086] 在本说明书的描述中,参考术语“一个实施例”、“示例”、“具体示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不一定指的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任何一个或多个实施例或示例中以合适的方式结合。

[0087] 以上公开的本发明优选实施例只是用于帮助阐述本发明。优选实施例并没有详尽叙述所有的细节,也不限制该发明仅为所述的具体实施方式。显然,根据本说明书的内容,可作很多的修改和变化。本说明书选取并具体描述这些实施例,是为了更好地解释本发明的原理和实际应用,从而使所属技术领域技术人员能很好地理解和利用本发明。本发明仅受权利要求书及其全部范围和等效物的限制。

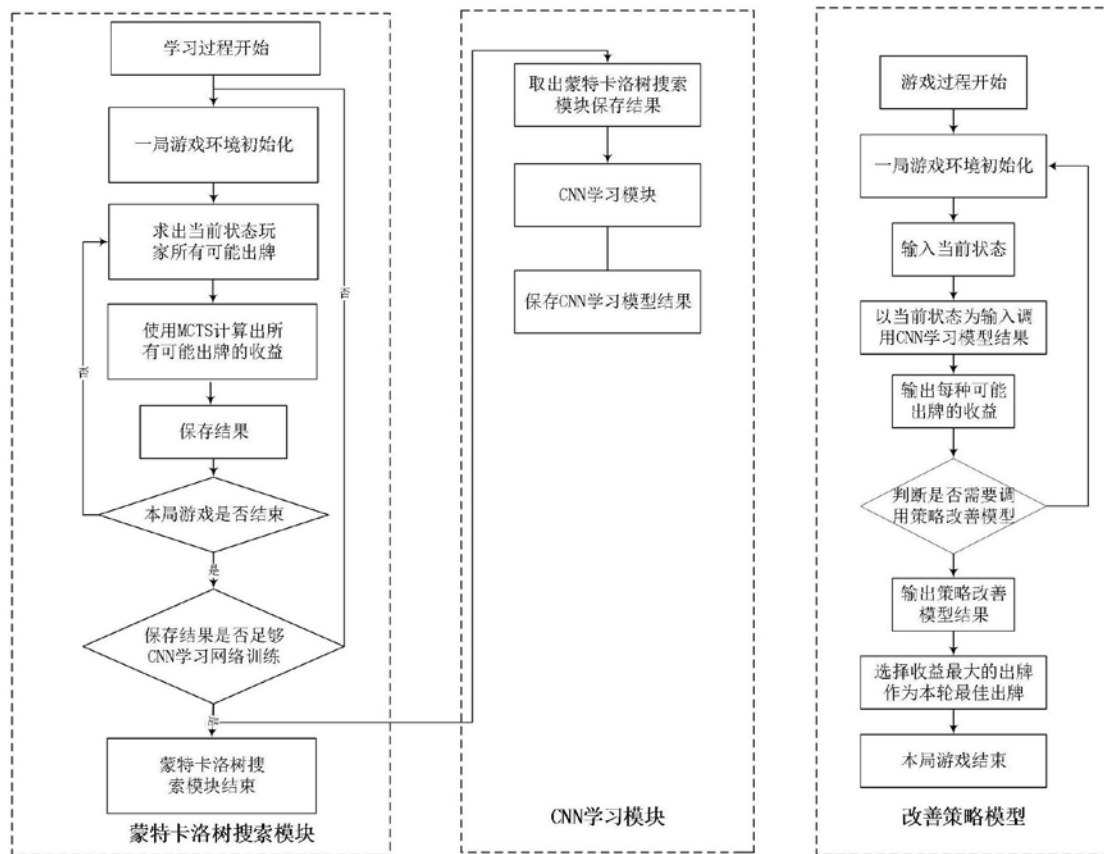


图1

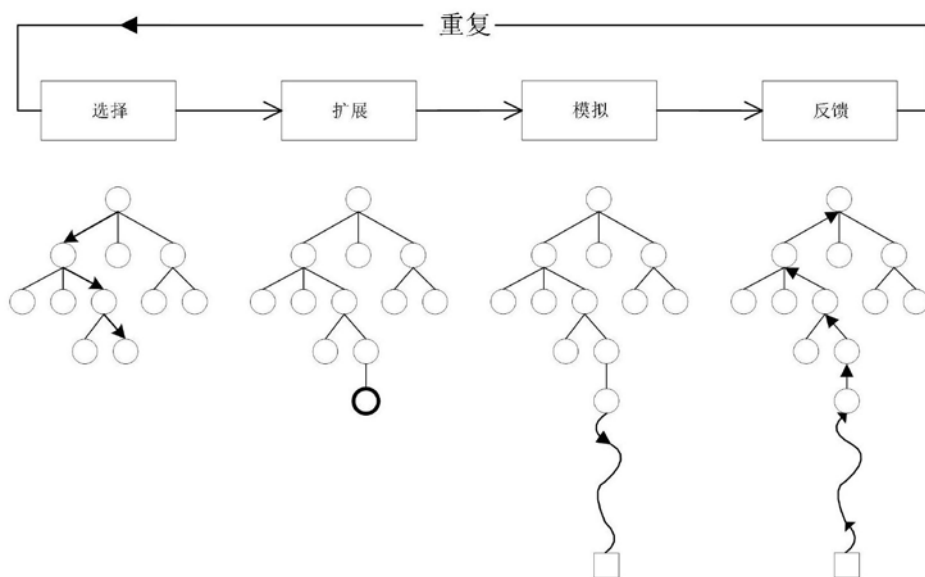


图2

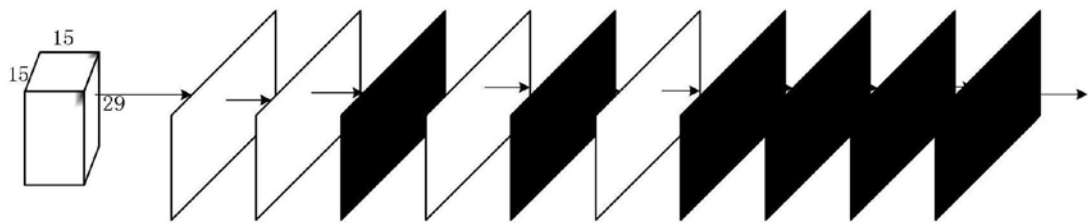


图3

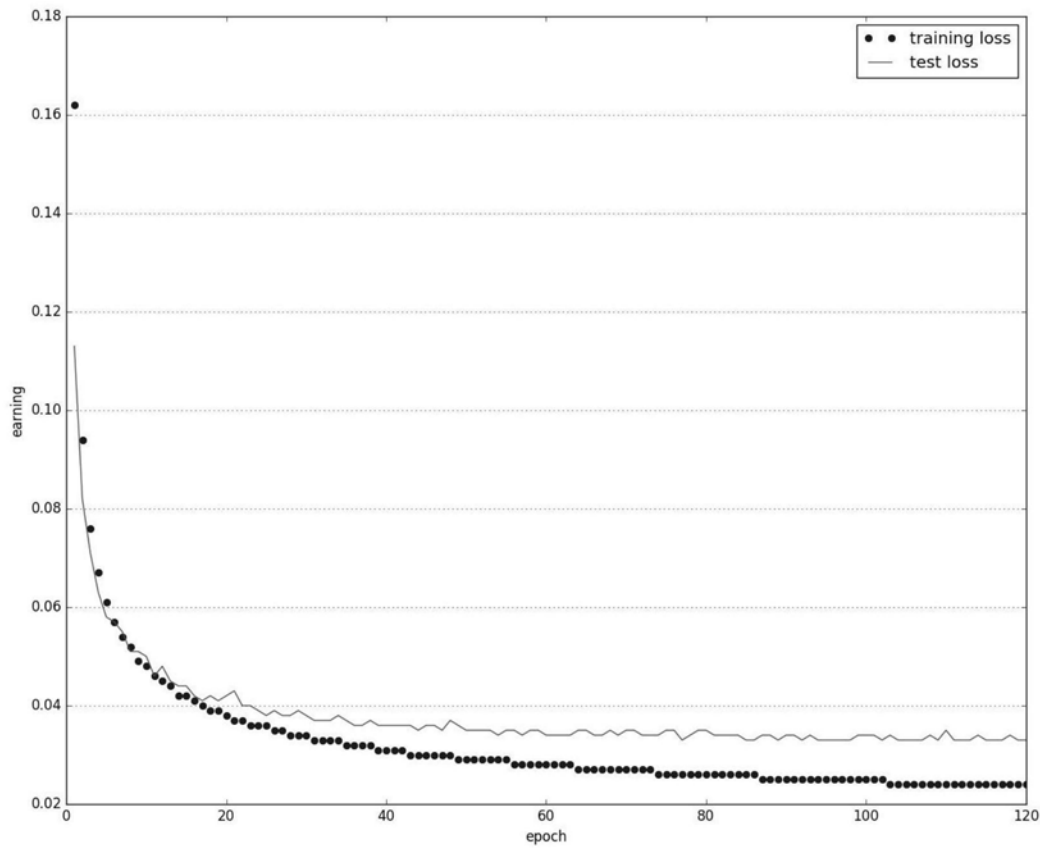


图4

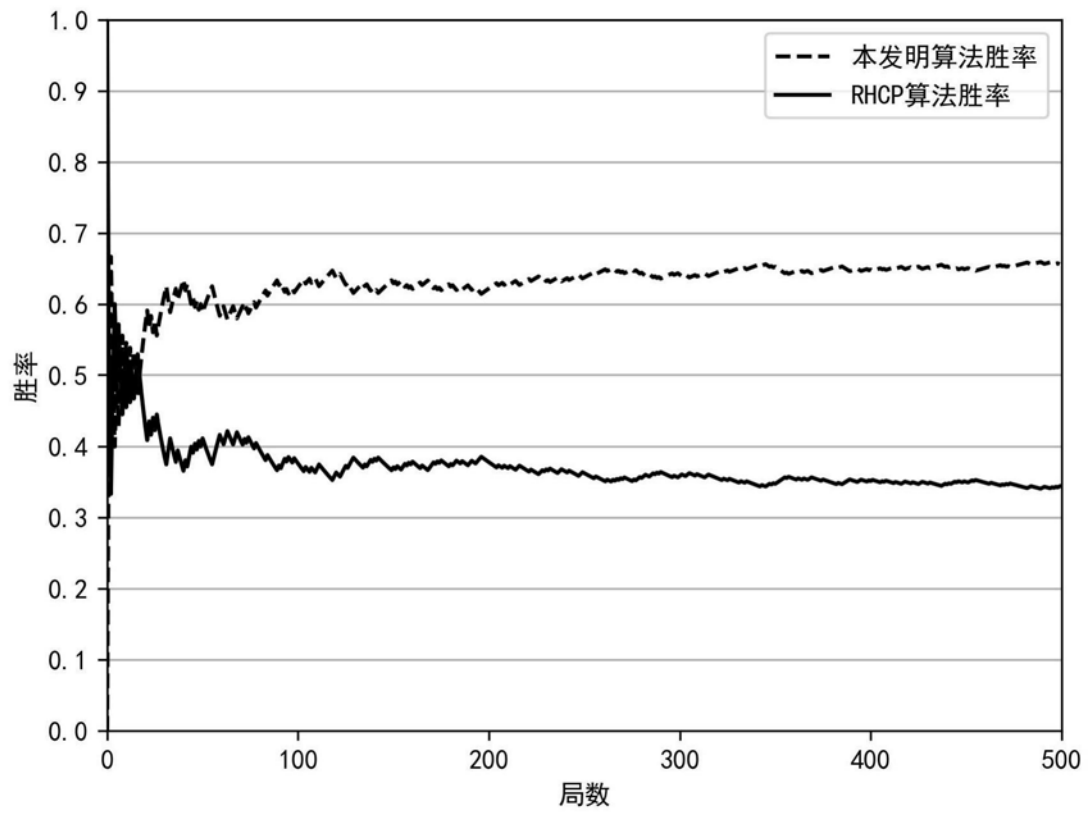


图5

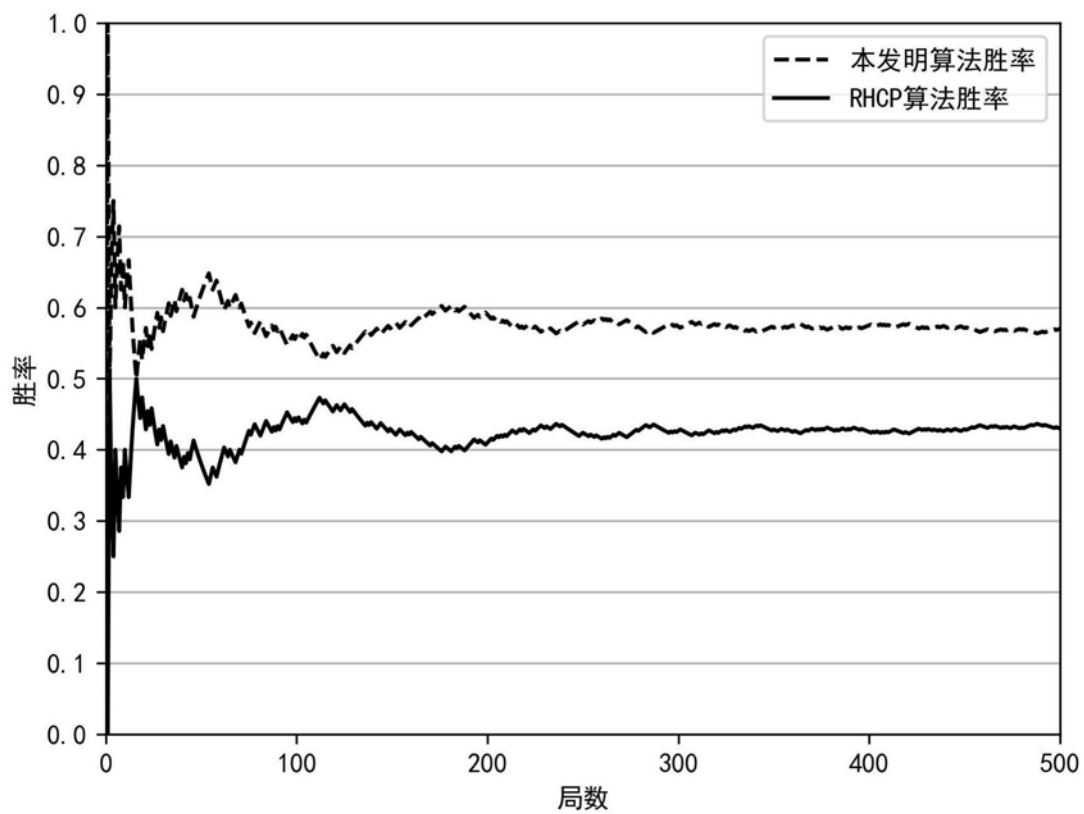


图6