(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111931910 A
(43)申请公布日 2020.11.13

(21)申请号 202010735083.4

(22)申请日 2020.07.28

(71)申请人 西交利物浦大学
地址 215121 江苏省苏州市工业园区独墅湖科教创新区仁爱路111号

(72)发明人 苏炯龙 蒋正雍 高源 高梓铭 胡奕

(74)专利代理机构 南京艾普利德知识产权代理事务所(特殊普通合伙) 32297
代理人 陆明耀 顾祥安

(51)Int.Cl.
G06N 3/04(2006.01)
G06N 3/08(2006.01)
G06Q 40/04(2012.01)
G06Q 40/06(2012.01)

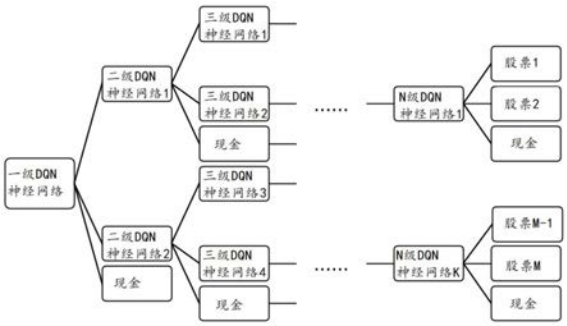
权利要求书3页 说明书9页 附图6页

(54)发明名称

基于强化学习和深度学习的投资方法及智能体

(57)摘要

本发明揭示了基于强化学习和深度学习的投资方法及智能体,方法通过构建多层式Deep Q-network模型,除底层的DQN神经网络外,高层的DQN神经网络用于将其所管理资产分配给其下一层的DQN神经网络及现金,底层DQN神经网络用于将其管理的资产分配给其管理的股票和现金,以使下一个交易周期该股票的收益最高;用训练完成后的模型获取下一交易周期初的资产分配权重,并调整资产在股票市场中的分配,从而得到最优投资策略。本方案采用分层式结构进行建模,一定程度上缩减了动作空间规模以及超参数数量,极大地提高了对动作空间的探索效率,并且简化了神经网络结构,降低了神经网路的训练难度,使神经网络能够有效地进行训练,并找出每个状态下的最优动作。



CN 111931910 A

1.基于强化学习和深度学习的投资方法,其特征在于:包括如下步骤:

S1,构建股票交易场景的多层式Deep Q-network模型;

所述多层式Deep Q-network模型包括多层DQN神经网络,所述DQN神经网络的输入为第n个交易周期的价格张量,除底层的DQN神经网络外,高层的DQN神经网络用于将其所管理资产分配给其下一层的DQN神经网络及现金,底层的每个DQN神经网络用于将其管理的资产分配给其管理的股票和现金,以使下一个交易周期所述股票的收益最高,每个DQN神经网络对应一个马尔科夫决策过程;

S2,训练所述多层式Deep Q-network模型使其参数最优化;

S3,加载训练好的多层式Deep Q-network模型参数,接收实时股票价格数据,通过模型获取下一交易周期初的资产分配权重,并根据资产分配权重调整资产在股票市场中的分配,从而得到最优投资策略。

2.根据权利要求1的基于强化学习和深度学习的投资方法,其特征在于:所述S1中,每个DQN神经网络的输入为第n个交易周期的价格张量,其通过如下方法得到:

S11,分别提取所要管理股票的前N天的开盘、收盘、最高、最低价格,形成四个(M*N)的矩阵,M为该马尔科夫决策所要管理的组合优化的股票的个数,对于前N天中非交易日的数据,用上一个交易日的开盘价,收盘价,最高价,最低价对相应指标分别进行填充;

S12,将S11中得到的四个矩阵分别除以上一交易日的收盘价,使每个矩阵都标准化;

S13,采用扩张因子α,对标准化后矩阵P_t^*进行如下操作:

$P_t=\alpha(P_t^*-1)$;

S14,将经过S13得到的四个矩阵组合成一个(M,N,4)维的价格张量,即为第n个交易周期的价格张量。

3.根据权利要求1所述的基于强化学习和深度学习的投资方法,其特征在于:所述S2中,每层DQN神经网络的训练过程如下:

S21,提取记忆批次$\{(S_{t_1},a_{t_1},r_{t_1},S_{t_1+1}),\ldots\ldots,(S_{t_n},a_{t_n},r_{t_n},S_{t_n+1})\}$;

S22,计算该批次中每个记忆所对应的目标Q值

$$Q_{target^*(i)}=Q_{target}(S_{t_i+1},argmax\left(Q_{eval}(S_{t_i+1},a)\right))$$

其中,$Q_{eval}(S_{t_i+1},a)$表示状态S_{t_i+1}对应的每个动作的Q值估计值;

$argmax\left(Q_{eval}(S_{t_i+1},a)\right)$表示最大估计Q值所对应的动作;

$Q_{target}(S_{t_i+1},argmax\left(Q_{eval}(S_{t_i+1},a)\right))$为目标值网络给出的状态$S_{t_i+1}$的Q值目标值;

S23,计算该批次中每个记忆所对应的估计Q值

$$Q_{eval^*(i)}=Q_{eval}\left(S_{t_i},argmax\left(Q_{eval}(S_{t_i},a)\right)\right);$$

S24,得到目标Q值向量和估计Q值向量

$$\boldsymbol{Q}_{target^*} = [Q_{target^*(1)}, Q_{target^*(2)}, \cdots, Q_{target^*(n)}]\ ;$$

$$\boldsymbol{Q}_{eval^*} = [Q_{eval^*(1)}, Q_{eval^*(2)}, \cdots, Q_{eval^*(n)}]\ ;$$

S25,计算真实Q值并得到真实Q值向量

$$Q_{real(i)} = r_{t_i} + \gamma Q_{target^*(i)}$$

$Q_{real}=[Q_{real(1)},Q_{real(2)},\cdots,Q_{real(n)}]$;

S26,对损失函数$l(\boldsymbol{\theta}) = ||\boldsymbol{Q}_{eval^*} - \boldsymbol{Q}_{real}||^2$使用梯度下降,从而找到最优参数θ*使该损失函数值达到最小,其中θ为估计值Q网络Q_{eval}的参数;

S27,将最优参数赋给目标值Q网络Q_{target};

S28,重复上述步骤直至损失函数收敛。

4.根据权利要求1-3任一所述的基于强化学习和深度学习的投资方法,其特征在于:所述多层式Deep Q-network模型至少包括顶层DQN神经网络及底层DQN神经网络,并通过顶层DQN神经网络及底层DQN神经网络共同制定最终投资策略。

5.根据权利要求4所述的基于强化学习和深度学习的投资方法,其特征在于:所述顶层DQN神经网络的结构及其预测Q值的过程如下:

S311,接收(M,N,4)维的价格张量;

S312,通过两层卷积层对输入的张量进行特征提取,形成一个(64*M*1)的张量,这里M为投资产品个数;

S313,将S312中获得的(64*M*1)张量,通过卷积神经网络的Flatten层,将多维数据转化为一维数据;

S314,在S313得到的一维数据中插入上一交易周期结束后的资产分配比形成一个新的((M*64+2)*1)向量;

S315,将S314得到的向量插入现金偏置,形成一个((M*64+3)*1)的向量;

S316,将S315形成的向量通过一层含有2046个神经元的全连接层,形成一个(1*1)的状态向量Q_s和一个$\left(\binom{2+T}{2},\ 1\right)$的动作向量$Q_a$,其中$\binom{2+T}{2}$为动作空间内不同动作的个数;

S317,通过S316中得到的2个向量及公式$Q=Q_s+(Q_a-E[Q_a])$计算在该状态下的Q值,其中$E[Q_a]$为动作向量中所有值的均值,并在训练模型的过程中不断更新每个S1中定义动作的Q值;

S318,最后选取Q值最大的动作作为下一交易周期初分配给底层DQN及现金的资产权重向量。

6.根据权利要求5所述的基于强化学习和深度学习的投资方法,其特征在于:所述312包括

S3121,通过一层卷积核规模为1*3的卷积层得到32个(M*5)的特征矩阵,其中选用Selu函数作为神经元的激活函数;

S3122,将S3121得到的32个特征矩阵输入到一层卷积核规模为1*5的卷积层输出一个(64*M*1)的张量,其中选Selu函数作为神经元的激活函数。

7.根据权利要求4所述的基于强化学习和深度学习的投资方法,其特征在于:所述底层DQN神经网络的结构及其预测Q值的过程如下:

S321,接收(2,N,4)维或(1,N,4)维的价格张量;

S322,通过与S312结构相同的两层卷积层对输入的价格张量进行特征提取,形成一个(64*2*1)的向量;

S323,将S322得到的向量中插入上一交易周期结束后的资产分配比形成一个新的(65*2*1)向量;

S324,将S323得到的向量经过一层卷积核为1*1,激活函数为Selu的卷积层,形成(128*2*1)的向量,并通过卷积神经网络的Flatten层,将多维数据转化为一维数据;

S325,将S324得到的向量插入现金偏置,形成一个((2*128+1)*1)的向量;

S326,将S325得到的向量通过一层含有2046个神经元的全连接层,转化为一个(1*1)的状态向量Q_s和一个$\binom{2+T}{2}, 1)$或$\binom{1+T}{1}, 1)$的动作向量Q_a,其中$\binom{2+T}{2}$、$\binom{1+T}{1}$分别为动作空间内不同动作的个数;

S327,通过S326中得到的2个向量,根据公式$Q = Q_s + (Q_a - E[Q_a])$计算在特定状态下的Q值,并在训练模型的过程中不断更新每个S1中定义动作所对应的Q值;

S328,最后选取Q值最大的动作作为下一交易周期初分配给该底层DQN所负责的股票及现金的资产权重向量。

8.根据权利要求1所述的基于强化学习和深度学习的投资方法,其特征在于:还包括S4,定期使用不同的数据对多层式Deep Q-network模型进行叠加训练及参数微调。

9.智能体,其特征在于:包括多层式Deep Q-network模型,所述多层式Deep Q-network模型包括多层DQN神经网络,所述DQN神经网络的输入为第n个交易周期的价格张量,除底层的DQN神经网络外,高层的DQN神经网络用于将其所管理资产分配给其下一层的DQN神经网络及现金,底层的每个DQN神经网络用于将其管理的资产分配给其管理的股票和现金,以使下一个交易周期所述股票的收益最高。

基于强化学习和深度学习的投资方法及智能体

技术领域

[0001] 本发明涉及机器学习中的深度学习、增强学习的技术领领域，尤其是基于强化学习和深度学习的投资方法及智能体。

背景技术

[0002] 随着人工智能技术的发展，强化学习算法已经被应用于金融领域。目前，通过搭建合适的交互环境，基于Q Learning的强化学习Deep Q-network(DQN)模型已经被初步地应用于资产管理，例如申请号为201810030006.1所示方法。

[0003] 但是由于资产管理中动作空间过于庞大，单个DQN无法对其充分探索，因此单个DQN模型的收益状况并不出色。

[0004] 在资产管理的DQN模型中，动作被定义为每一交易周期初所确定的资产分配权重，在此定义下，再规定最小权重单位便可得到离散化的动作空间。然而，如果最小权重单位太小或者资产数量太多，就会导致资产权重向量(动作)数量过多，从而导致智能体在随机动作探索阶段无法对每个动作进行充分探索，并且使Deep Q-network的全连接层出现大量神经元。这样一来，不仅对于动作空间的探索效率很低，而且神经网络的训练也会变得非常困难。

[0005] 因此，如果使用单个DQN模型进行资产管理，通常需要限制资产个数，并且忽略手续费，以此来减少权重向量的个数。这严重影响了该模型的应用价值和泛化能力。因此本发明提出一种基于分层式Deep Q-network算法的股票交易方法及系统来应用于繁琐复杂的金融市场。

发明内容

[0006] 本发明的目的就是为了解决现有技术中存在的上述问题，提供一种基于强化学习和深度学习的投资方法及智能体。

[0007] 本发明的目的通过以下技术方案来实现：

[0008] 基于强化学习和深度学习的投资方法，包括如下步骤：

[0009] S1，构建股票交易场景的多层式Deep Q-network模型；

[0010] 通过深度神经网络构建智能代理，所述智能代理与使用股票开盘、收盘、最高、最低价格的时间序列数据构建的环境进行交互，

[0011] 智能代理的状态空间$\{S_n\}$定义为第n个交易周期的价格张量以及上一个交易周期资金分配比例情况所构成的二维数组；

[0012] 智能代理的动作a定义为进行交易后的资产分配比例，动作空间的规模定义为平均分的T份资产与M+1个投资品(包含现金)的随机组合$\binom{M+T}{M}$；

[0013] 所述多层式Deep Q-network模型包括多层DQN神经网络，除底层的DQN神经网络外，高层的DQN神经网络用于将其所管理资产分配给其下一层的DQN神经网络及现金，底层的每个DQN神经网络用于将其管理的资产分配给其管理的股票和现金，以使下一个交易周

期所述股票的收益最高。

[0014] 每个DQN神经网络对应一个马尔科夫决策过程，每个马尔科夫决策过程的动作空间大小为$\binom{2+T}{2}$或$\binom{1+T}{1}$，每个DQN神经网络的输入为第n个交易周期的价格张量，每个DQN神经网络对应的马尔科夫决策的奖励r如下：

[0015] r＝所负责资产经过一个交易周期的资产总额/上一时期所负责资产的资产总额；

[0016] S2，训练所述多层式Deep Q-network模型使其参数最优化；

[0017] S3，加载训练好的多层式Deep Q-network模型参数，接收实时股票价格数据，通过模型获取下一交易周期初的资产分配权重，并根据资产分配权重调整资产在股票市场中的分配，从而得到最优投资策略。

[0018] 优选的，所述基于强化学习和深度学习的投资方法中，所述S1中，所述第n个交易周期的价格张量通过如下方法得到：

[0019] S11，分别提取所要管理股票的前N天的开盘、收盘、最高、最低价格，形成四个(M*N)的矩阵，M为该马尔科夫决策所要管理的组合优化的股票的个数，对于前N天中非交易日的数据，用上一个交易日的，收盘价，开盘价，最高价，最低价对相应指标分别进行填充；

[0020] S12，将S11中得到的四个矩阵分别除以上一交易日的收盘价，使每个矩阵都标准化；

[0021] S13，采用扩张因子α，对标准化后矩阵P_t^*进行如下操作：

[0022] $P_t=\alpha(P_t^*-1)$；

[0023] S14，将经过S13得到的四个矩阵组合成一个(M,N,4)维的价格张量，即为第n个交易周期的价格张量。

[0024] 优选的，所述基于强化学习和深度学习的投资方法中，所述S2中，每层DQN神经网络的训练过程如下：

[0025] S21，提取记忆批次$\{(S_{t_1},a_{t_1},r_{t_1},S_{t_1+1}),\ldots\ldots,(S_{t_n},a_{t_n},r_{t_n},S_{t_n+1})\}$；

[0026] S22，计算该批次中每个记忆所对应的目标Q值

[0027] $$Q_{target^*(i)}=Q_{target}(S_{t_i+1},argmax\left(Q_{eval}(S_{t_i+1},a)\right))$$

[0028] 其中，$Q_{eval}(S_{t_i+1},a)$表示状态S_{t_i+1}对应的每个动作的Q值的估计值；

[0029] $argmax\left(Q_{eval}(S_{t_i+1},a)\right)$表示最大估计Q值所对应的动作；

[0030] $Q_{target}(S_{t_i+1},argmax\left(Q_{eval}(S_{t_i+1},a)\right))$为目标值网络给出的状态$S_{t_i+1}$的Q值目；

[0031] S23，计算该批次中每个记忆所对应的估计Q值

[0032] $$Q_{eval^*(i)}=Q_{eval}\left(S_{t_i},argmax\left(Q_{eval}(S_{t_i},a)\right)\right);$$

[0033] S24，得到目标Q值向量和估计Q值向量

[0034] $$\boldsymbol{Q}_{target^*}=[Q_{target^*(1)},Q_{target^*(2)},\ldots,Q_{target^*(n)}]\ ;$$

[0035] $\boldsymbol{Q}_{eval^*} = [Q_{eval^*(1)}, Q_{eval^*(2)}, \cdots, Q_{eval^*(n)}]$ ；

[0036] S25,计算真实Q值并得到真实Q值向量

[0037] $Q_{real(i)} = r_{t_i} + \gamma Q_{target^*(i)}$

[0038] Q_{real}＝[$Q_{real(1)}$,$Q_{real(2)}$,…,$Q_{real(n)}$]；

[0039] S26,对损失函数$\mathrm{l}(\boldsymbol{\theta}) = \ ||\boldsymbol{Q}_{eval^*} - \boldsymbol{Q}_{real}||^2$使用梯度下降,从而找到最优参数θ*使该损失函数值达到最小,其中θ为估计值Q网络Q_{eval}的参数；

[0040] S27,将最优参数赋给目标值Q网络Q_{target}；

[0041] S28,重复上述步骤直至损失函数收敛。

[0042] 优选的,所述基于强化学习和深度学习的投资方法中,所述多层式Deep Q-network模型至少包括顶层DQN神经网络及底层DQN神经网络,并通过顶层DQN神经网络及底层DQN神经网络共同制定最终投资策略。

[0043] 优选的,所述基于强化学习和深度学习的投资方法中,在S3中,所述顶层DQN神经网络的结构及其预测Q值的过程如下：

[0044] S311,接收(M,N,4)维的价格张量；

[0045] S312,通过两层卷积层对输入的张量进行特征提取,形成一个(64*M*1)的张量,M为投资产品个数；

[0046] S313,将S312中获得的(64*M*1)张量,通过卷积神经网络的Flatten层,将多维数据转化为一维数据；

[0047] S314,在S313得到的一维数据中插入上一交易周期结束后的资产分配比形成一个新的((M*64+2)*1)向量；

[0048] S315,将S314得到的向量插入现金偏置,形成一个((M*64+3)*1)的向量；

[0049] S316,将S315形成的向量通过一层含有2046个神经元的全连接层,形成一个(1*1)的状态向量Q_s和一个($\binom{2+T}{2}$,1)的动作向量Q_a,其中$\binom{2+T}{2}$为动作空间内不同动作的个数；

[0050] S317,通过S316中得到的2个向量及公式Q＝Q_s+(Q_a-E[Q_a])计算在该状态下的Q值,其中E[Q_a]为动作向量中所有值的均值,并在训练模型的过程中不断更新每个S1中定义动作的Q值；

[0051] S318,最后选取Q值最大的动作作为下一交易周期初分配给底层DQN及现金的资产权重向量。

[0052] 优选的,所述基于强化学习和深度学习的投资方法中,所述312包括

[0053] S3121,通过一层卷积核规模为1*3的卷积层得到32个(M*5)的特征矩阵,其中选用Selu函数作为神经元的激活函数；

[0054] S3122,将S3121得到的32个特征矩阵输入到一层卷积核规模为1*5的卷积层输出一个(64*M*1)的张量,其中选用Selu函数作为神经元的激活函数。

[0055] 优选的,所述基于强化学习和深度学习的投资方法中,所述底层DQN神经网络的结构及其预测Q值的过程如下：

[0056] S321,接收(2,N,4)维或(1,N,4)维的价格张量；

[0057] S322,通过与S312结构相同的两层卷积层对输入的价格张量进行特征提取,形成一个(64*2*1)的向量；

[0058] S323,将S322得到的向量中插入上一交易周期结束后的资产分配比形成一个新的(65*2*1)向量;

[0059] S324,将S323得到的向量经过一层卷积核为1*1,激活函数为Selu的卷积层,形成(128*2*1)的向量,并通过卷积神经网络的Flatten层,将多维数据转化为一维数据;

[0060] S325,将S324得到的向量插入现金偏置,形成一个((2*128+1)*1)的向量;

[0061] S326,将S325得到的向量通过一层含有2046个神经元的全连接层,转化为一个(1*1)的状态向量Q_s和一个($\binom{2+T}{2}$,1)或($\binom{1+T}{1}$,1)的动作向量Q_a,其中$\binom{2+T}{2}$、$\binom{1+T}{1}$分别为动作空间内不同动作的个数;

[0062] S327,通过S326中得到的2个向量,根据公式$Q=Q_s+(Q_a-E[Q_a])$计算在特定状态下的Q值,并在训练模型的过程中不断更新每个S1中定义动作所对应的Q值;

[0063] S328,最后选取Q值最大的动作作为下一交易周期初分配给该底层DQN所负责的股票及现金的资产权重向量。

[0064] 优选的,所述基于强化学习和深度学习的投资方法中,还包括S4,定期使用不同的数据对多层式Deep Q-network模型进行叠加训练及参数微调。

[0065] 智能体,包括多层式Deep Q-network模型,所述多层式Deep Q-network模型包括多层DQN神经网络,所述DQN神经网络的输入为第n个交易周期的价格张量,除底层的DQN神经网络外,高层的DQN神经网络用于将其所管理资产分配给其下一层的DQN神经网络及现金,底层的每个DQN神经网络用于将其管理的资产分配给其管理的股票和现金,以使下一个交易周期所述股票的收益最高。

[0066] 本发明技术方案的优点主要体现在:

[0067] 本方案采用分层式结构进行建模,一定程度上缩减了动作空间规模以及超参数数量,使神经网络能够有效地进行训练,并找出每个状态下的最优动作,极大地提高了探索效率,简化了神经网络结构,降低了神经网路的训练难度。

附图说明

[0068] 图1是本发明的分层式Deep Q-network模型的一般结构示意图;

[0069] 图2是本发明的分层式Deep Q-network模型包括两层DQN神经网络的结构示意图;

[0070] 图3是顶层DQN神经网络的结构及工作原理示意图;

[0071] 图4是底层DQN神经网络的结构及工作原理示意图;

[0072] 图5-图7是2013、2016、2017三个时间段的测试数据集进行不同模型测试的结果比对图。

具体实施方式

[0073] 本发明的目的、优点和特点,将通过下面优选实施例的非限制性说明进行图示和解释。这些实施例仅是应用本发明技术方案的典型范例,凡采取等同替换或者等效变换而形成的技术方案,均落在本发明要求保护的范围之内。

[0074] 在方案的描述中,需要说明的是,术语“中心”、“上”、“下”、“左”、“右”、“前”、“后”、“竖直”、“水平”、“内”、“外”等指示的方位或位置关系为基于附图所示的方位或位置关系,

仅是为了便于描述和简化描述，而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作，因此不能理解为对本发明的限制。此外，术语“第一”、“第二”、“第三”仅用于描述目的，而不能理解为指示或暗示相对重要性。并且，在方案的描述中，以操作人员为参照，靠近操作者的方向为近端，远离操作者的方向为远端。

[0075] 下面结合附图对本发明揭示的基于强化学习和深度学习的投资方法进行阐述，其包括如下步骤：

[0076] S1，构建股票交易场景的多层式Deep Q-network模型；

[0077] 具体的，通过深度神经网络构建智能代理，所述智能代理与使用股票

[0078] 开

[0079] 盘、收盘、最高、最低价格的时间序列数据构建的环境进行交互，环境会产生状态转移和即时回报，通过状态转移和即时回报的数据，训练深度神经网络，再次采取动作，依照上述过程循环，使智能代理每次采取动作的累计折扣即时回报最大化。

[0080] 其中，智能代理的状态空间$\{S_n\}$定义为第n个交易周期的价格张量以及上一个交易周期资金分配比例情况所构成的二维数组。

[0081] 所述第n个交易周期的价格张量作为深度神经网络的每个DQN神经网络的输入，通过如下方法得到：

[0082] S11，分别提取所要管理股票的前N天的开盘、收盘、最高、最低价格，形成四个(M*N)的矩阵，M为该马尔科夫决策所要管理的组合优化的股票的个数，对于前N天中非交易日的数据，用上一个交易日的，收盘价，开盘价，最高价，最低价对相应指标分别进行填充。

[0083] S12，将S11中得到的四个矩阵分别除以上一交易日的收盘价，使每个矩阵都标准化。

[0084] S13，采用扩张因子α，对标准化后矩阵P_t^*进行如下操作：

[0085] $P_t=\alpha(P_t^*-1)$。

[0086] S14，将S13得到的四个矩阵组合成一个(M,N,4)维的价格张量，即为第n个交易周期的价格张量。

[0087] 智能代理的动作a定义为进行交易后的资产分配比例，动作空间的规模定义为平均分的T份资产与M+1个投资品(包含现金)的随机组合$\binom{M+T}{M}$。

[0088] 为了减小DQN神经网络的动作空间，因此采用如附图1所示的多层式Deep Q-network模型，这种分层式Deep Q-nework模型为一种多智能体结合的模型，其主要特点在于有非常明确的分层结构。该分层结构将模型中的多个DQN神经网络分为不同等级(level)，位于更高等级的DQN神经网络将所持有的资产分配给更低等级的DQN神经网络，再由最后一级的DQN神经网络将其所持有的资产分配给其管理的股票和现金。

[0089] 即所述多层式Deep Q-network模型包括多层DQN神经网络，除底层的DQN神经网络外，高层的DQN神经网络用于将其所管理资产分配给其下一层的DQN神经网络及现金，底层的每个DQN神经网络用于将其管理的资产分配给其管理的股票和现金，以使下一个交易周期所述股票的收益最高。

[0090] 具体的，如附图2所示，所述多层式Deep Q-network模型至少包括

[0091] 顶层DQN神经网络，其作用在于将其所负责的资产分配给下一级DQN神经网络，其

具体结构如附图3所示；

[0092] 底层DQN神经网络，其作用是将其管理的资产(由上一级DQN神经网络分配得到)分配到其管理的股票和现金中，其具体结构如附图4所示。

[0093] 当然，在实际应用中，由于要管理的股票个数较多，在顶层DQN神经网络与底层DQN神经网络之间可能需要加入中间层DQN神经网络。中间层DQN神经网络的层数根据要管理的股票总数进行设计，每层的每个DQN神经网络的结构及构建方法与顶层DQN神经网络一致，此处不作赘述。

[0094] 此时，每个DQN神经网络即对应一个马尔科夫决策过程，每个马尔科夫决策过程的动作空间缩小到$\binom{2+T}{2}$或$\binom{1+T}{1}$。

[0095] 每个DQN神经网络对应的马尔科夫决策的奖励r如下：

[0096] r＝所负责资产经过一个交易周期的资产总额/上一时期所负责资产的资产总额。

[0097] S2，通过训练集数据训练所述多层式Deep Q-network模型使其参数最优化，具体训练时，每层DQN神经网络的训练过程如下：

[0098] S21，提取记忆批次$\{(S_{t_1}, a_{t_1}, r_{t_1}, S_{t_1+1}), \ldots\ldots, (S_{t_n}, a_{t_n}, r_{t_n}, S_{t_n+1})\}$，所述记忆批次由经验池(experience replay)随机抽取。

[0099] S22，计算该批次中每个记忆所对应的目标Q值

[0100] $$Q_{target^*(i)} = Q_{target}(S_{t_i+1}, argmax\left(Q_{eval}(S_{t_i+1}, a)\right))$$

[0101] 其中，$Q_{eval}(S_{t_i+1}, a)$表示状态S_{t_i+1}对应的每个动作的Q值的估计值；

[0102] $argmax\left(Q_{eval}(S_{t_i+1}, a)\right)$表示最大估计Q值所对应的动作；

[0103] $Q_{target}(S_{t_i+1}, argmax\left(Q_{eval}(S_{t_i+1}, a)\right))$为目标值网络给出的状态$S_{t_i+1}$的Q值目。

[0104] S23，计算该批次中每个记忆所对应的估计Q值

[0105] $$Q_{eval^*(i)} = Q_{eval}\left(S_{t_i}, argmax\left(Q_{eval}(S_{t_i}, a)\right)\right)。$$

[0106] S24，得到目标Q值向量和估计Q值向量

[0107] $$\boldsymbol{Q}_{target^*} = [Q_{target^*(1)}, Q_{target^*(2)}, \ldots, Q_{target^*(n)}]；$$

[0108] $$\boldsymbol{Q}_{eval^*} = [Q_{eval^*(1)}, Q_{eval^*(2)}, \ldots, Q_{eval^*(n)}]。$$

[0109] S25，计算真实Q值并得到真实Q值向量

[0110] $$Q_{real(i)} = r_{t_i} + \gamma Q_{target^*(i)}$$

[0111] Q_{real}＝[$Q_{real(1)}$，$Q_{real(2)}$，…，$Q_{real(n)}$]。

[0112] S26，对损失函数$\mathrm{I}(\boldsymbol{\theta}) = ||\boldsymbol{Q}_{eval^*} - \boldsymbol{Q}_{real}||^2$使用梯度下降，从而找到最优参数$\theta^*$使该损失函数值达到最小，其中θ为估计值Q网络$Q_{eval}$的参数。

[0113] S27，将最优参数赋给目标值Q网络Q_{target}。

[0114] S28，重复上述步骤直至损失函数收敛，具体判断收敛的方式可根据损失函数的图

像,起初有明显下降趋势,然后保持在一个稳定水平即为收敛。

[0115] 训练好的模型能够直接用于股票交易,即根据不同的实时股票价格,输出对应的资金分配情况。

[0116] S3,加载训练好的多层式Deep Q-network模型参数,接收实时股票价格数据,通过模型获取下一交易周期初的资产分配权重,并根据资产分配权重调整资产在股票市场中的分配,从而得到最优投资策略。

[0117] 此处,为了方便说明,所述多层式Deep Q-network模型以两层为例,其包括顶层DQN神经网络及底层DQN神经网络,其中,顶层DQN神经网络负责决策如何将总资产有效地分配给现金、底层DQN神经网络1及底层DQN神经网络2;底层DQN神经网络1及底层DQN神经网络2负责决策如何将顶层DQN神经网络分配给其的资产额最有效地分配给现金和该底层DQN神经网络所负责的两支股票,使下一个交易周期该两个股票的收益最高。当然,每个底层DQN神经网络也可以仅管理一支股票。

[0118] 如附图3所示,所述顶层DQN神经网络预测Q值的过程如下:

[0119] S311,接收(M,N,4)维的价格张量。

[0120] S312,通过两层卷积层对输入的张量进行特征提取,形成一个(64*M*1)的张量,这里M为投资产品个数;该步骤具体包括

[0121] S3121,通过一层卷积核规模为1*3的卷积层得到32个(M*5)的特征矩阵,其中选用Selu函数作为神经元的激活函数;

[0122] S3122,将S3121得到的32个特征矩阵输入到一层卷积核规模为1*5的卷积层输出一个(64*M*1)的张量,其中选用Selu函数作为神经元的激活函数。

[0123] S313,将S312中获得的(64*M*1)张量,通过卷积神经网络的Flatten层,将多维数据转化为一维数据。

[0124] S314,在S313得到的一维数据中插入上一交易周期结束后的资产分配比形成一个新的((M*64+2)*1)向量。

[0125] S315,将S314得到的向量插入现金偏置,形成一个((M*64+3)*1)的向量。

[0126] S316,将S315形成的向量通过一层含有2046个神经元的全连接层,形成一个(1*1)的状态向量Q_s和一个($\binom{2+T}{2}$,1)的动作向量Q_a,其中$\binom{2+T}{2}$为动作空间内不同动作的个数。

[0127] S317,通过S316中得到的2个向量及公式$Q=Q_s+(Q_a-E[Q_a])$计算在该状态下的Q值,其中$E[Q_a]$为动作向量中所有值的均值,并在训练模型的过程中不断更新每个S1中定义动作的Q值。

[0128] S318,最后选取Q值最大的动作作为下一交易周期初分配给底层DQN及现金的资产权重向量。

[0129] 如附图4所示,所述底层DQN神经网络预测Q值的过程如下:

[0130] S321,接收(2,N,4)维或(1,N,4)维的价格张量,其中,当底层的一个DQN神经网络管理两支股票时,接收(2,N,4)的价格张量;当底层的一个DQN神经网络管理一支股票时,接收(1,N,4)维的价格张量。

[0131] S322,通过与S312结构相同的两层卷积层对输入的价格张量进行特征提取,形成一个(64*2*1)的向量,具体过程同S3121-S3122。

[0132] S323,将S322得到的向量中插入上一交易周期结束后的资产分配比形成一个新的(65*2*1)向量。

[0133] S324,将S323得到的向量经过一层卷积核为1*1,激活函数为Selu的卷积层,形成(128*2*1)的向量,并通过卷积神经网络的Flatten层,将多维数据转化为一维数据。

[0134] S325,将S324得到的向量插入现金偏置,形成一个((2*128+1)*1)的向量。

[0135] S326,将S325得到的向量通过一层含有2046个神经元的全连接层,转化为一个(1*1)的状态向量Q_s和一个($\binom{2+T}{2}$,1)或($\binom{1+T}{1}$,1)的动作向量Q_a,其中$\binom{2+T}{2}$、$\binom{1+T}{1}$分别为动作空间内不同动作的个数。

[0136] S327,通过S326中得到的2个向量,根据公式$Q=Q_s+(Q_a-E[Q_a])$计算在特定状态下的Q值,并在训练模型的过程中不断更新每个S1中定义动作所对应的Q值。

[0137] S328,最后选取Q值最大的动作作为下一交易周期初分配给该底层DQN所负责的股票及现金的资产权重向量。

[0138] 最终投资策略是由顶层DQN神经网络和底层DQN神经网络共同制定,即顶层DQN负责分配资产到底层DQN,底层DQN负责将被分配到的资产投入到股票市场中。

[0139] 在模型训练完成后,投入使用前,可以采用测试数据集进行模型的性能检测,具体的,以四支相关性较低的A股股票为例构成测试数据集,四支股票代码分别为:600260、600261、600262和600266。具体是通过雅虎金融下载四支股票的时间序列数据,分别提取所选股票的前N天的开盘、收盘、最高、最低价格,形成四个4*N的价格矩阵。将得到的四组数据合并上无风险资产(现金),处理成规模为四个5*N的包含五种投资产品的价格矩阵。对于前N天中非交易日的数据,用上一个交易日的开盘价,收盘价,最高价,最低价相对应的指标分别进行填充。并且通过四个矩阵分别除以上一交易日的收盘价,使每个矩阵都标准化。此后,采用扩张因子α,使标准化后矩阵的列的数值之间差异更加显著。实际将2013/1/14-2013/12/19、2016/1/14-2016/12/19和2017/1/13-2017/12/18的股票时间序列数据分别设定为测试数据集。

[0140] 将测试数据集输入本方案的模型(H-DQN)后所得的结果和传统资产管理方法结果的对比呈现如附图5-附图7所示,其中涉及的传统方法如下:

[0141] • Robust Median Reversion(RMR)

[0142] • Uniform Buy and Hold(BAH)

[0143] • Universal Portfolios(UP)

[0144] • Exponential Gradient(EG)

[0145] • Online Newton Step(ONS)

[0146] • Aniticor(ANTICOR)

[0147] • Passive Aggressive Mean Reversion(PAMR)

[0148] • Online Moving Average Reversion(OLMAR)

[0149] • Confidence Weighted Mean Reversion(CWMR)

[0150] • Single DQN(N-DQN)。

[0151] 从比对图可以看出,本方案的方案相对其他方法,其获得的收益水平最佳,取得了更好的效果。

[0152] S4,针对复杂多变的股票市场,模型需要进行增量式的训练,因此,定期使用不同特征的价格时间序列数据对多层式Deep Q-network模型进行叠加训练及参数微调,这也是一种迁移学习的过程,使得模型更加完善和健壮,使模型拥有更好的扩展性和鲁棒性,此处具体的训练过程与上述S21-S28的过程相同,在此不作赘述。

[0153] 本方案进一步揭示了一种智能体,包括上述多层式Deep Q-network模型,所述多层式Deep Q-network模型包括多层DQN神经网络,所述DQN神经网络的输入为第n个交易周期的价格张量,除底层的DQN神经网络外,高层的DQN神经网络用于将其所管理资产分配给其下一层的DQN神经网络及现金,底层的每个DQN神经网络用于将其管理的资产分配给其管理的股票和现金,以使下一个交易周期所述股票的收益最高。当然,所述智能体还包括其他各种已知智能体所具有的通用结构,例如数据采集模块、执行器等,此处为已知技术,不作赘述。

[0154] 本发明尚有多种实施方式,凡采用等同变换或者等效变换而形成的所有技术方案,均落在本发明的保护范围之内。

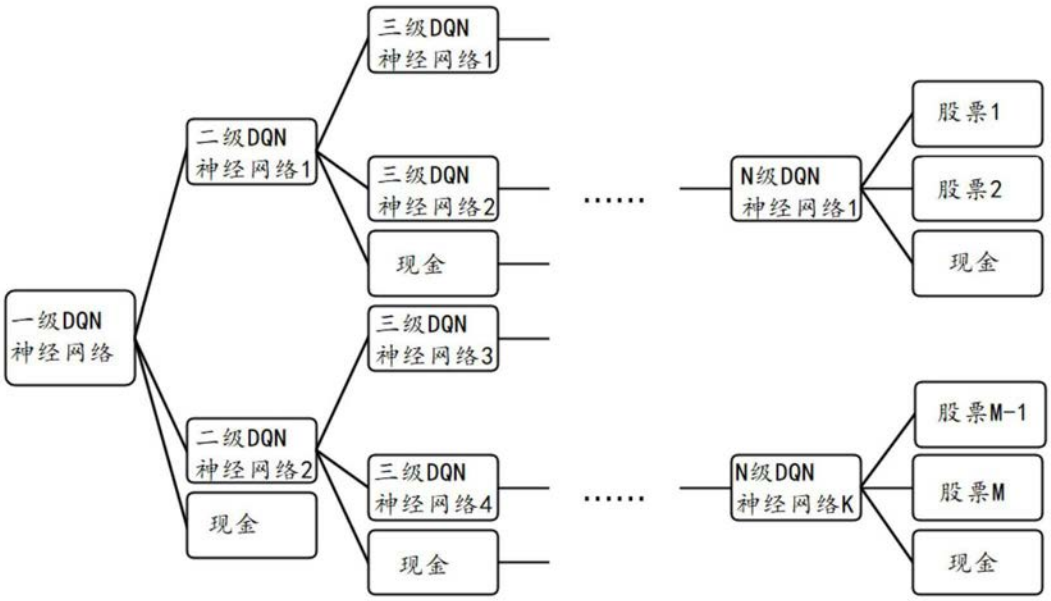

一级DQN
神经网络
二级DQN
神经网络1
二级DQN
神经网络2
现金
三级DQN
神经网络1
三级DQN
神经网络2
现金
三级DQN
神经网络3
三级DQN
神经网络4
现金
……
……
N级DQN
神经网络1
N级DQN
神经网络K
股票1
股票2
现金
股票M-1
股票M
现金

图1

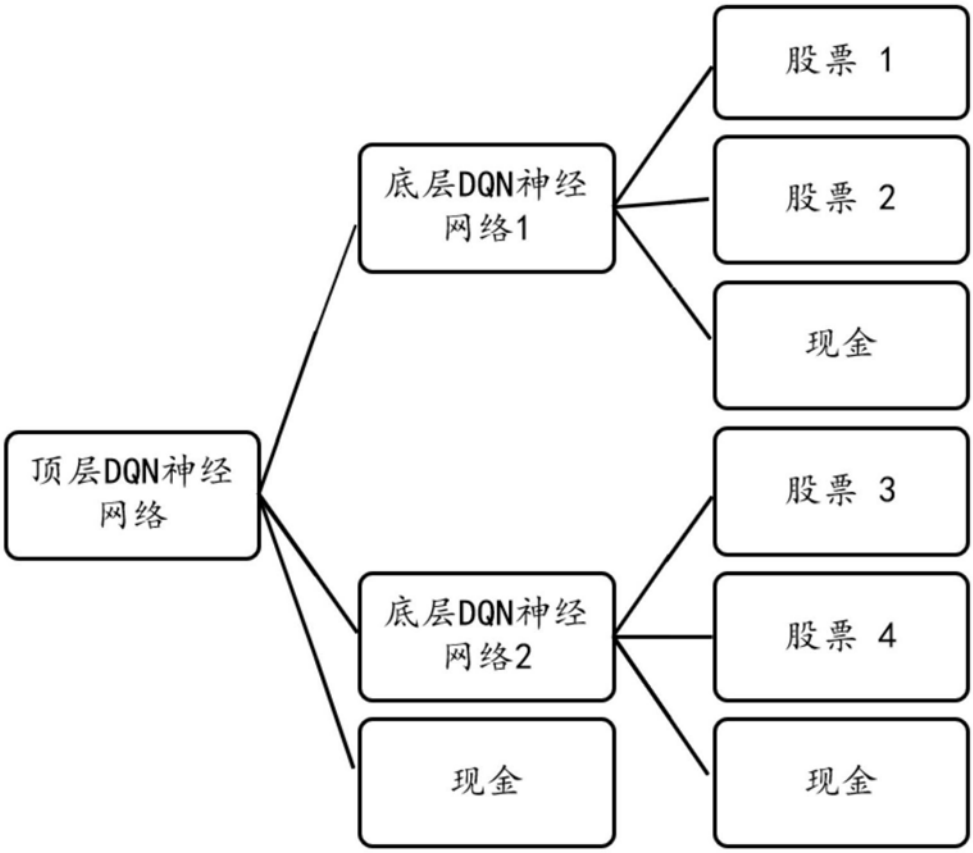

顶层DQN神经
网络
底层DQN神经
网络1
底层DQN神经
网络2
现金
股票 1
股票 2
现金
股票 3
股票 4
现金

图2

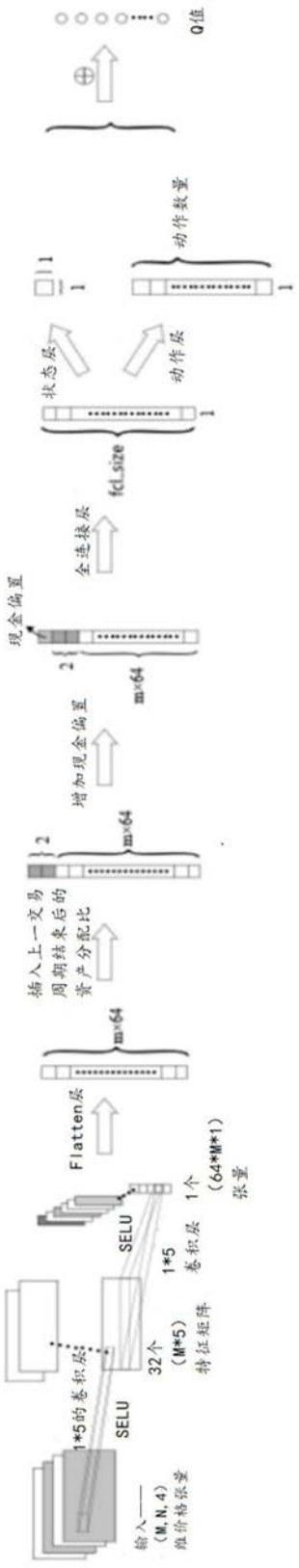

输入——
(M,N,4)
股价格张量
1*5的卷积层
SELU
32个
(M*5)
特征矩阵
1*5
卷积层
SELU
1个
(64*M*1)
张量
Flatten层
m×64
插入上一交易
周期结束后的
资产分配比
2
m×64
增加现金偏置
现金偏置
2
m×64
全连接层
fcl_size
1
状态层
动作层
1
1
1
动作数量
1
Q值

图3

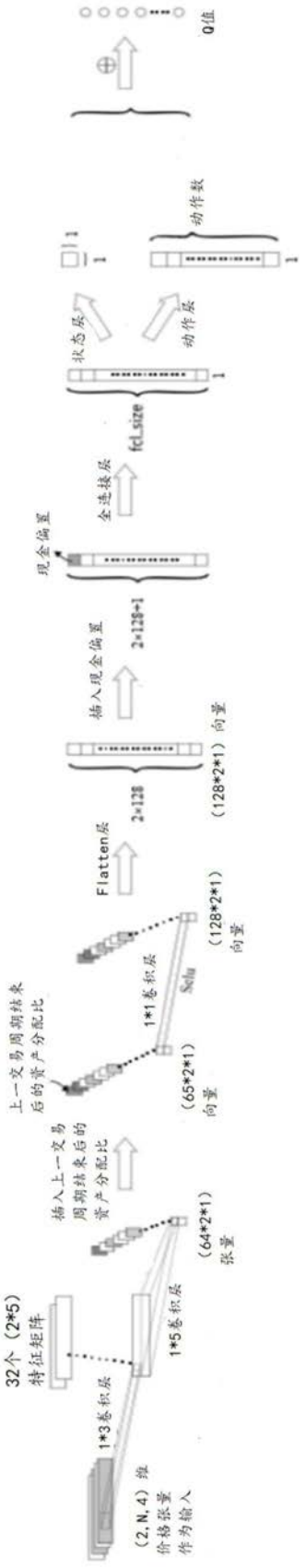

(2,N,4)维
价格张量
作为输入
1*3卷积层
32个（2*5）
特征矩阵
1*5卷积层
(64*2*1)
张量
插入上一交易
周期结束后的
资产分配比
上一交易周期结束
后的资产分配比
(65*2*1)
向量
1*1卷积层
Selu
(128*2*1)
向量
Flatten层
2*128
(128*2*1)向量
插入现金偏置
2*128+1
现金偏置
全连接层
fcl_size
状态层
动作层
1
1
动作数
1
Q值

图4

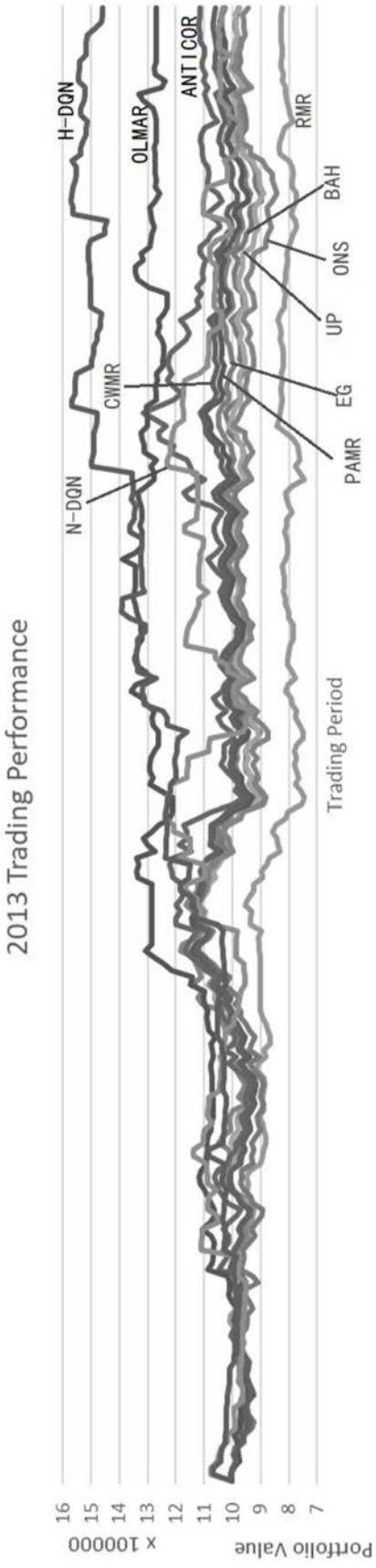

2013 Trading Performance
Trading Period
Portfolio Value
x 100000
16
15
14
13
12
11
10
9
8
7
H-DQN
OLMAR
ANTICOR
RMR
BAH
ONS
UP
EG
CWMR
PAMR
N-DQN

图5

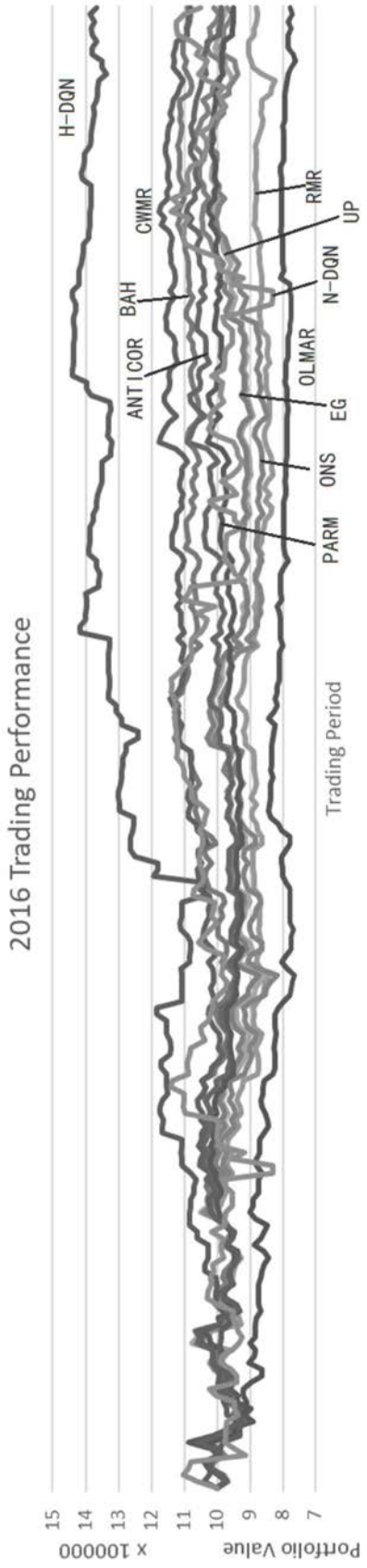

2016 Trading Performance
Portfolio Value
x 100000
15
14
13
12
11
10
9
8
7
Trading Period
H-DQN
CWMR
BAH
ANTICOR
OLMAR
PARM
ONS
EG
N-DQN
UP
RMR

图6

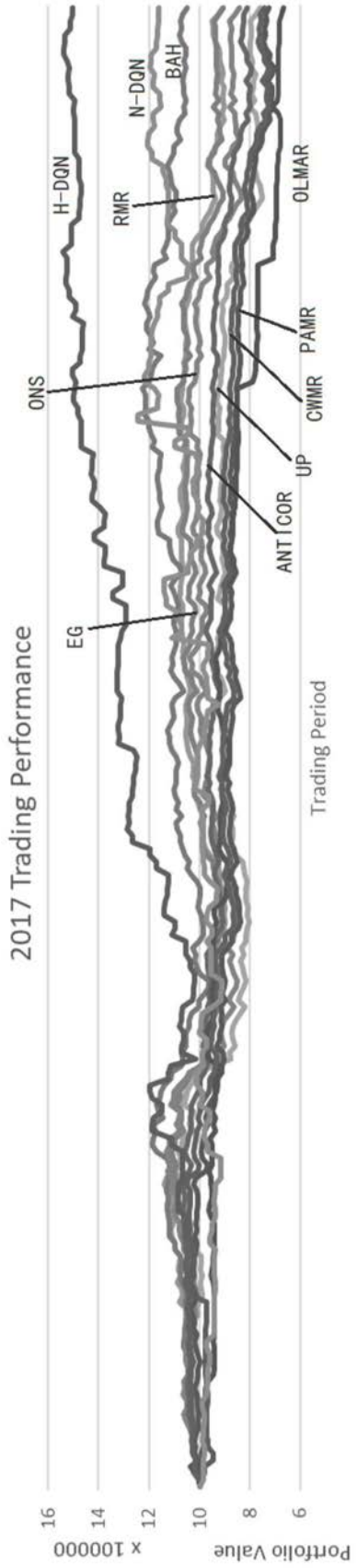

2017 Trading Performance
Portfolio Value
x 100000
16
14
12
10
8
6
Trading Period
H-DQN
N-DQN
BAH
RMR
OLMAR
PAMR
CWMR
UP
COR
ANTI
ONS
EG

图7