(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 108596335 A
(43)申请公布日 2018.09.28

(21)申请号 201810362557.8

(22)申请日 2018.04.20

(71)申请人 浙江大学
地址 310058 浙江省杭州市西湖区余杭塘路866号

(72)发明人 张寅 杨璞 胡滨

(74)专利代理机构 杭州求是专利事务所有限公司 33200
代理人 傅朝栋 张法高

(51)Int.Cl.
G06N 3/08(2006.01)
G06Q 10/06(2012.01)

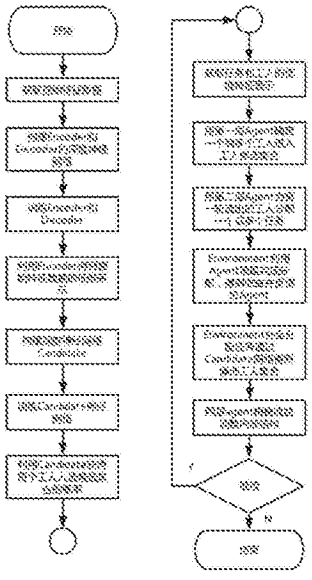
权利要求书2页 说明书5页 附图2页

(54)发明名称

一种基于深度强化学习的自适应众包方法

(57)摘要

本发明公开了一种基于深度强化学习的自适应众包方法。方法具体为:1)首先从众包系统中采样需要分配的任务和候选的众包工人;2)通过深度学习方法获得待分配任务和候选工人的低维特征表示;3)通过强化学习方法确定任务分配策略;4)众包系统根据分配策略分配任务,根据任务完成结果评估本次分配获得的收益,将该收益反馈给强化学习方法,更新强化学习参数;5)从1)开始继续下一轮的任务分配。和现有技术相比,本发明结合了深度强化学习方法,系统地对任务分配问题进行建模,针对不同任务本身的特征选择合适的众包工人,形成了自适应的智能众包方法,创造性地提升了众包的工作效率和效果。



CN 108596335 A

1.一种基于深度强化学习的自适应众包方法,其特征在于,步骤如下:

S1.首先从众包系统中采样需要分配的众包任务和众包工人的信息;

S2.通过深度学习方法获得待分配任务和工人的低维特征表示,具体包括以下子步骤:

S21.获取原始特征数据,包括众包任务的原始特征和众包工人的原始特征;

S22.构建深度神经网络,包括Encoder和Decoder两部分,其中Encoder的输入为原始特征数据,输出为原始特征的低维表示;Decoder的输入为Encoder所得的低维表示,输出为该低维表示的解析结果,即原始特征数据的近似表达;

S23.一同训练Encoder和Decoder,输入设定为原始特征数据,损失函数设定为原始特征数据与Decoder最终输出的距离,训练使得Encoder-Decoder的输出逼近原始特征数据;

S24.使用训练好的Encoder,输入原始特征数据后获得原始特征数据的低维表示;

S3.通过深度学习方法获得每个工人入选候选集合的概率,遴选候选工人,具体包括以下子步骤:

S31.构建深度神经网络Candidate,输入为工人的低维特征表示,输出为该工人入选候选集合的概率;

S32.训练Candidate,输入设定为工人的低维特征表示、工人得到任务后完成任务的概率,损失函数设定为工人完成任务概率和Candidate最终输出的距离,训练使得Candidate的输出逼近工人完成任务的概率,即工人任务完成率越高,工人入选候选集合概率越高;

S33.使用训练好的Candidate,获得每个待分配工人入选候选集合的概率,并依概率将工人选入候选集合;

S4.通过强化学习方法确定任务分配策略,完成本轮任务执行,具体包括以下子步骤:

S41.将待分配任务和候选工人的低维特征作为强化学习Agent第一层的输入,第一层Agent根据其内部的深度神经网络确定一个到多个工人;

S42.根据第一层Agent确定的工人,选取Agent第二层并输入待分配的任务,Agent第二层根据其内部的深度神经网络确定一个到多个任务进行分配,即确定任务分配策略,交由Environment执行;

S43.得到Environment分配策略后立即完成分配,工人执行完分配的任务后计算本轮任务分配获得的收益;

S5.根据上一轮执行结果,优化强化学习参数并更新工人的原始特征数据,并执行步骤S2-S4,具体包括以下子步骤:

S51.根据上一轮任务执行结果,将Environment计算的收益反馈给强化学习两层Agent,两层Agent根据获得的收益反馈,调整内部的深度神经网络,提高选择高收益策略的概率,降低选择低收益策略的概率;

S52.根据上一轮任务执行结果,更新工人的原始特征数据;

S53.Environment保留上一轮未分配的任务,通过随机采样补全待分配任务,获得新一轮的待分配任务;并再次执行步骤S2获得新一轮的候选工人集合;

S54.将新一轮的待分配任务和候选工人集合的原始特征作为输入,再次执行所述步骤S3和S4;

S6.不断重复步骤S5直到众包任务完成。

2.根据权利要求1所述的一种基于深度强化学习的自适应众包方法,其特征在于,步骤

S1中，所述众包任务的原始特征包括任务分类标签、任务文本内容、预估困难程度；所述众包工人的原始特征包括别、年龄、完成任务时间分布、历史总分配任务数、历史总完成任务数、各类任务分配和完成数。

3.根据权利要求1所述的一种基于深度强化学习的自适应众包方法，其特征在于，步骤S4中，所述的Agent第一层神经网络通过计算每个工人的预期收益，选择一到多个预期收效最高的工人进行分配，并根据每轮任务收益的反馈，调整计算工人预期收益相关的参数。

4.根据权利要求1所述的一种基于深度强化学习的自适应众包方法，其特征在于，步骤S4中，所述的Agent第二层神经网络通过计算每个任务的预期收益，选择一个到多个预期收益最高的任务分配给工人，并根据每轮任务收益的反馈，调整计算任务预期收益相关的参数。

5.根据权利要求1所述的一种基于深度强化学习的自适应众包方法，其特征在于，步骤S4中，所述Agent的第一、二层的各个单元采用不同的强化学习方法，所述强化学习方法包括Q-learning、DQN、DPG、DDPG；第二层Agent的每个单元对应一个工人，单元数量根据工人数量自适应变化。

6.根据权利要求1所述的一种基于深度强化学习的自适应众包方法，其特征在于，步骤S5中，所述的收益反馈可根据众包需求针对性设定：若众包设定的目标是尽可能多地完成任务，则收益反馈的内容为任务最终的完成数量；若众包设定的目标是尽可能正确地完成任务，则反馈为完成任务的准确率；若众包设定的目标是同时兼顾上述两种目标，则反馈为任务最终完成数量与完成任务准确率的加权求和。

一种基于深度强化学习的自适应众包方法

技术领域

[0001] 本发明涉及深度强化学习方法在众包系统上的应用，尤其涉及众包系统中工人遴选、任务分配的技术方法。

背景技术

[0002] 随着互联网的快速发展及信息全球化的推进，众包模式应运而生。众包是互联网带来的新的生产组织形式，改变了传统的解决方案，是一种分布式解决问题的方式，即利用互联网将相关工作分解并分配出去，化整为零。通过给予参与用户适当的奖励，将空闲生产力利用起来。对于政府和非盈利性组织而言，众包被认为是一种有潜力的问题解决机制。

[0003] 众包在数据标注、图书电子化、知识图谱构建等方面都有着广泛的应用。在数据标注方面，海量非结构化数据需要人为标注转化为结构化数据，包括有监督深度学习在内的一系列方法都需要大量结构化数据作为支撑。而这些数据标注任务难以在短时间内由少数人完成。在图书电子化领域，数字图书馆的蓬勃发展使得人们可以通过互联网访问海量的图书资源，节能环保，但现存的扫描版电子书需要大量的人力物力转换为文本数据。虽然目前的OCR技术已较为成熟，但仍有大量的识别错误需要人为修正。除此之外，知识图谱构建也面临着类似问题。虽然知识图谱能够挖掘、分析、构建、绘制、显示知识及其相互关系，为学科研究提供切实的、有价值的参考，但知识图谱的构建过程中命名实体识别、实体关系抽取等任务均需要人工的参与。面对诸如此类的困境，众包技术的使用可以大大提高工作效率、降低投入成本。

[0004] 在众包技术的应用中，任务的具体分配会很大程度影响到生产效率。一份不够完善的分配方案很有可能导致冗余工作的产生，增大成本，降低产出；反之，一份完善的分配方案能更大程度发挥众包技术的优势，提高空闲生产力的利用率。本发明重点对任务分配过程进行建模，将任务集合和工人集合的特点(即任务和工人的原始特征数据)和众包应用的任务目标相结合，用深度强化学习的方法获得一份完善的分配方案。

发明内容

[0005] 为了解决现有技术中存在的问题，本发明提供了一种基于深度强化学习的自适应众包方法。

[0006] 本发明结合深度学习和强化学习方法确定任务分配策略。对于某个具体目标的众包应用，本发明首先通过深度学习方法遴选众包工人，之后再运用强化学习方法确定具体任务分配，并根据最终任务的完成情况与目标的契合程度的反馈更新强化学习算法参数，优化分配策略。通过结合深度学习与强化学习，本发明不仅保证了任务分配方案契合众包应用的最终目标，保障了众包的质量，同时还完成了方法结构的分层，使得任务分配更具灵活性。

[0007] 为实现上述目的，本发明的技术方案为：

[0008] 基于深度强化学习的自适应众包方法，其步骤如下：

[0009] S1.首先从众包系统中采样需要分配的众包任务和众包工人的信息；

[0010] S2.通过深度学习方法获得待分配任务和工人的低维特征表示，具体包括以下子步骤：

[0011] S21.获取原始特征数据，包括众包任务的原始特征和众包工人的原始特征；

[0012] S22.构建深度神经网络，包括Encoder和Decoder两部分，其中Encoder的输入为原始特征数据，输出为原始特征的低维表示；Decoder的输入为Encoder所得的低维表示，输出为该低维表示的解析结果，即原始特征数据的近似表达；

[0013] S23.一同训练Encoder和Decoder，输入设定为原始特征数据，损失函数设定为原始特征数据与Decoder最终输出的距离，训练使得Encoder-Decoder的输出逼近原始特征数据；

[0014] S24.使用训练好的Encoder，输入原始特征数据后获得原始特征数据的低维表示；

[0015] S3.通过深度学习方法获得每个工人入选候选集合的概率，遴选候选工人，具体包括以下子步骤：

[0016] S31.构建深度神经网络Candidate，输入为工人的低维特征表示，输出为该工人入选候选集合的概率；

[0017] S32.训练Candidate，输入设定为工人的低维特征表示、工人得到任务后完成任务的概率，损失函数设定为工人完成任务概率和Candidate最终输出的距离，训练使得Candidate的输出逼近工人完成任务的概率，即工人任务完成率越高，工人入选候选集合概率越高；

[0018] S33.使用训练好的Candidate，获得每个待分配工人入选候选集合的概率，并依概率将工人选入候选集合；

[0019] S4.通过强化学习方法确定任务分配策略，完成本轮任务执行，具体包括以下子步骤：

[0020] S41.将待分配任务和候选工人的低维特征作为强化学习Agent第一层的输入，第一层Agent根据其内部的深度神经网络确定一个到多个工人；

[0021] S42.根据第一层Agent确定的工人，选取Agent第二层并输入待分配的任务，Agent第二层根据其内部的深度神经网络确定一个到多个任务进行分配，即确定任务分配策略，交由Environment执行；

[0022] S43.得到Environment分配策略后立即完成分配，工人执行完分配的任务后计算本轮任务分配获得的收益；

[0023] S5.根据上一轮执行结果，优化强化学习参数并更新工人的原始特征数据，并执行步骤S2-S4，具体包括以下子步骤：

[0024] S51.根据上一轮任务执行结果，将Environment计算的收益反馈给强化学习两层Agent，两层Agent根据获得的收益反馈，调整内部的深度神经网络，提高选择高收益策略的概率，降低选择低收益策略的概率；

[0025] S52.根据上一轮任务执行结果，更新工人的原始特征数据；

[0026] S53.Environment保留上一轮未分配的任务，通过随机采样补全待分配任务，获得新一轮的待分配任务；并再次执行步骤S2获得新一轮的候选工人集合；

[0027] S54.将新一轮的待分配任务和候选工人集合的原始特征作为输入，再次执行所述

步骤S3和S4；

[0028] S6.不断重复步骤S5直到众包任务完成。

[0029] 作为优选，所述众包任务的原始特征包括任务分类标签、任务文本内容、预估困难程度；所述众包工人的原始特征包括别、年龄、完成任务时间分布、历史总分配任务数、历史总完成任务数、各类任务分配和完成数。

[0030] 作为优选，步骤S4中，所述的Agent第一层神经网络通过计算每个工人的预期收益，选择一到多个预期收效最高的工人进行分配，并根据每轮任务收益的反馈，调整计算工人预期收益相关的参数。

[0031] 作为优选，步骤S4中，所述的Agent第二层神经网络通过计算每个任务的预期收益，选择一个到多个预期收益最高的任务分配给工人，并根据每轮任务收益的反馈，调整计算任务预期收益相关的参数。

[0032] 作为优选，步骤S4中，所述Agent的第一、二层的各个单元采用不同的强化学习方法，所述强化学习方法包括Q-learning、DQN、DPG、DDPG；第二层Agent的每个单元对应一个工人，单元数量根据工人数量自适应变化。

[0033] 作为优选，步骤S5中，所述的收益反馈可根据众包需求针对性设定：若众包设定的目标是尽可能多地完成任务，则收益反馈的内容为任务最终的完成数量；若众包设定的目标是尽可能正确地完成任务，也就是既快又好地完成任务，则反馈为完成任务的准确率；若众包设定的目标是同时兼顾上述两种目标，则反馈为任务最终完成数量与完成任务准确率的加权求和。

[0034] 和现有技术相比，本发明结合了深度强化学习方法，系统地对任务分配问题进行建模，针对不同任务本身的特征选择合适的众包工人，形成独具一格的自适应众包技术，创造性地提升了众包的工作效率和效果。

附图说明

[0035] 图1为基于深度强化学习的自适应众包方法的流程图；

[0036] 图2为基于深度强化学习的自适应众包方法模型图；

[0037] 图3为DQN模型图。

具体实施方式

[0038] 为了使本发明的目的、技术方案及优点更加清楚明白，以下结合附图对本发明进行进一步详细说明。

[0039] 参考图1，为本发明的基于深度强化学习的自适应众包方法的实施流程。基于深度强化学习的自适应众包方法包括以下步骤：

[0040] S1.首先从众包系统中采样需要分配的众包任务和众包工人的信息；

[0041] 本步骤中，所述众包任务的原始特征包括任务分类标签、任务文本内容、预估困难程度；所述众包工人的原始特征包括别、年龄、完成任务时间分布、历史总分配任务数、历史总完成任务数、各类任务分配和完成数。

[0042] S2.通过深度学习方法获得待分配任务和工人的低维特征表示，具体包括以下子步骤：

[0043] S21.获取原始特征数据，包括众包任务的原始特征和众包工人的原始特征；

[0044] S22.构建深度神经网络，包括Encoder和Decoder两部分，其中Encoder的输入为原始特征数据，输出为原始特征的低维表示；Decoder的输入为Encoder所得的低维表示，输出为该低维表示的解析结果，即原始特征数据的近似表达；

[0045] S23.一同训练Encoder和Decoder，输入设定为原始特征数据，损失函数设定为原始特征数据与Decoder最终输出的距离，训练使得Encoder-Decoder的输出尽量逼近原始特征数据；

[0046] S24.使用训练好的Encoder，输入原始特征数据后获得原始特征数据的低维表示；

[0047] S3.通过深度学习方法获得每个工人入选候选集合的概率，遴选候选工人，具体包括以下子步骤：

[0048] S31.构建深度神经网络Candidate，输入为工人的低维特征表示，输出为该工人入选候选集合的概率；

[0049] S32.训练Candidate，输入设定为工人的低维特征表示、工人得到任务后完成任务的概率，损失函数设定为工人完成任务概率和Candidate最终输出的距离，训练使得Candidate的输出尽量逼近工人完成任务的概率，即工人任务完成率越高，工人入选候选集合概率越高；

[0050] S33.使用训练好的Candidate，获得每个待分配工人入选候选集合的概率，并依概率将工人选入候选集合。

[0051] S4.通过强化学习方法确定任务分配策略，完成本轮任务执行，如图2所示，具体包括以下子步骤：

[0052] S41.将待分配任务和候选工人的低维特征作为强化学习Agent第一层的输入，第一层Agent根据其内部的深度神经网络确定一个到多个工人；

[0053] 本步骤中，所述的Agent第一层神经网络通过计算每个工人的预期收益，选择一到多个预期收效最高的工人进行分配，并根据每轮任务收益的反馈，调整计算工人预期收益相关的参数。

[0054] S42.根据第一层Agent确定的工人，选取Agent第二层并输入待分配的任务，Agent第二层根据其内部的深度神经网络确定一个到多个任务进行分配，即确定任务分配策略，交由Environment执行；

[0055] 本步骤中，所述的Agent第二层神经网络通过计算每个任务的预期收益，选择一个到多个预期收益最高的任务分配给工人，并根据每轮任务收益的反馈，调整计算任务预期收益相关的参数。

[0056] S43.得到Environment分配策略后立即完成分配，工人执行完分配的任务后计算本轮任务分配获得的收益；

[0057] 本步骤，所述Agent的第一、二层的各个单元可采用不同的强化学习方法，所述强化学习方法包括Q-learning、DQN、DPG、DDPG；第二层Agent的每个单元对应一个工人，单元数量根据工人数量自适应变化。其中，DQN模型图如图3所示；以Q-learning为例，其每个单元内部的深度神经网络可视为一个函数Q(s,a)，s为当前的状态输入(state)，a为当前的选择输入(action)，即评估当前输入下，每一种选择a的价值。其损失函数计算如下：

[0058] $$L(\theta) = E[\left(r + \gamma \max_{a'} Q(s', a') - Q(s, a)\right)^2]$$

[0059] 其中E()为期望函数,r为本次选择所获得的收益,γ为长期收益的折扣因子,s'为下一轮的状态输入,a'为下一轮的选择输入。这使得Q(s,a)能够不断地逼近输入s下,做出选择a的长期累计收益。在最终实施选择时,根据每个选择a的Q(s,a)值的高低,依概率确定一个到多个选择即可。

[0060] S5.根据上一轮执行结果,优化强化学习参数并更新工人的原始特征数据,并重复执行步骤S2-S4,具体包括以下子步骤:

[0061] S51根据上一轮任务执行结果,将Environment计算的收益反馈给强化学习两层Agent,两层Agent根据获得的收益反馈,调整内部的深度神经网络,提高选择高收益策略的概率,降低选择低收益策略的概率;

[0062] 本步骤中,所述的收益反馈可根据众包需求针对性设定:若众包设定的目标是尽可能多地完成任务,则收益反馈的内容为任务最终的完成数量;若众包设定的目标是尽可能正确地完成任务,则反馈为完成任务的准确率;若众包设定的目标是兼顾上述两种目标,既快又好地完成任务,则反馈为任务最终完成数量与完成任务准确率的加权求和。

[0063] S52根据上一轮任务执行结果,更新工人的原始特征数据;

[0064] S53 Environment保留上一轮未分配的任务,通过随机采样补全待分配任务,获得新一轮的待分配任务;并再次执行步骤S2获得新一轮的候选工人集合;

[0065] S54将新一轮的待分配任务和候选工人集合的原始特征作为输入,再次执行步骤S3、S4;

[0066] S6.不断重复步骤S5直到众包任务完成。

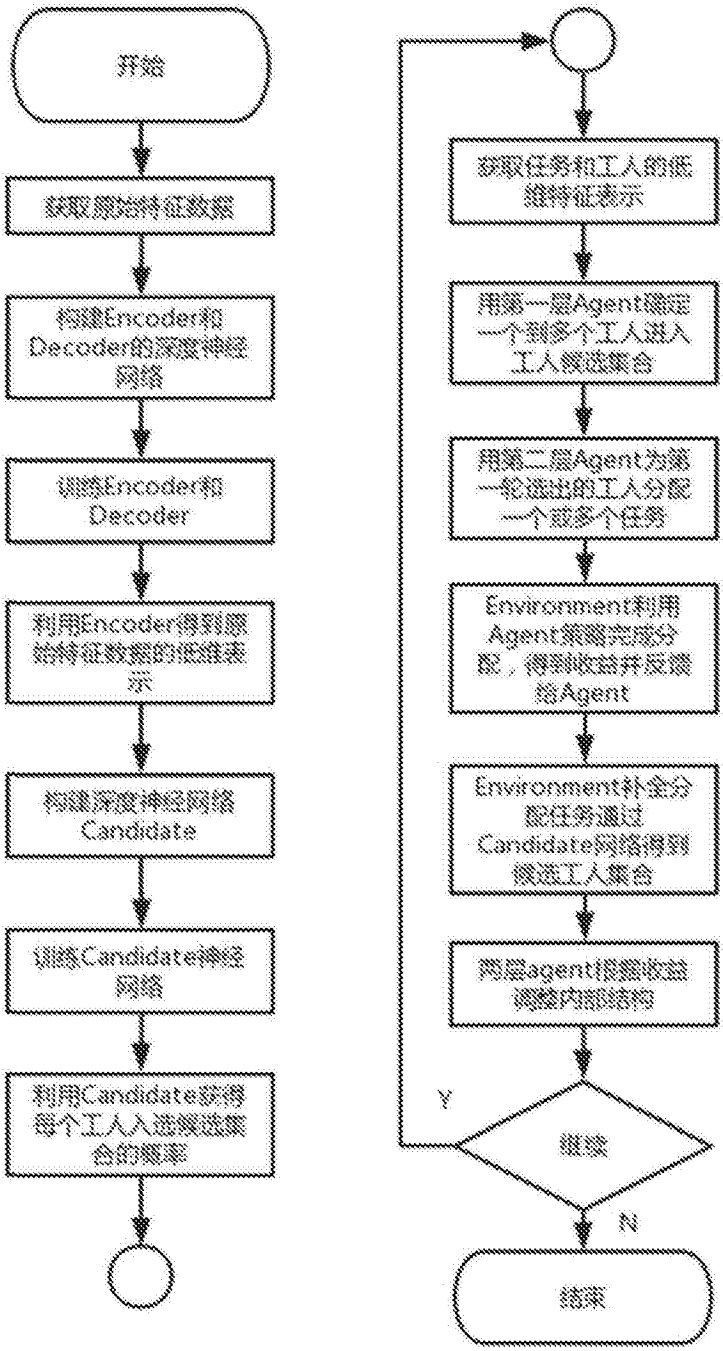

开始
获取原始特征数据
构建Encoder和Decoder的深度神经网络
训练Encoder和Decoder
利用Encoder得到原始特征数据的低维表示
构建深度神经网络Candidate
训练Candidate神经网络
利用Candidate获得每个工人入选候选集合的概率
获取任务和工人的低维特征表示
用第一层Agent确定一个到多个工人进入工人候选集合
用第二层Agent为第一轮选出的工人分配一个或多个任务
Environment利用Agent策略完成分配，得到收益并反馈给Agent
Environment补全分配任务通过Candidate网络得到候选工人集合
两层agent根据收益调整内部结构
Y
继续
N
结束

图1

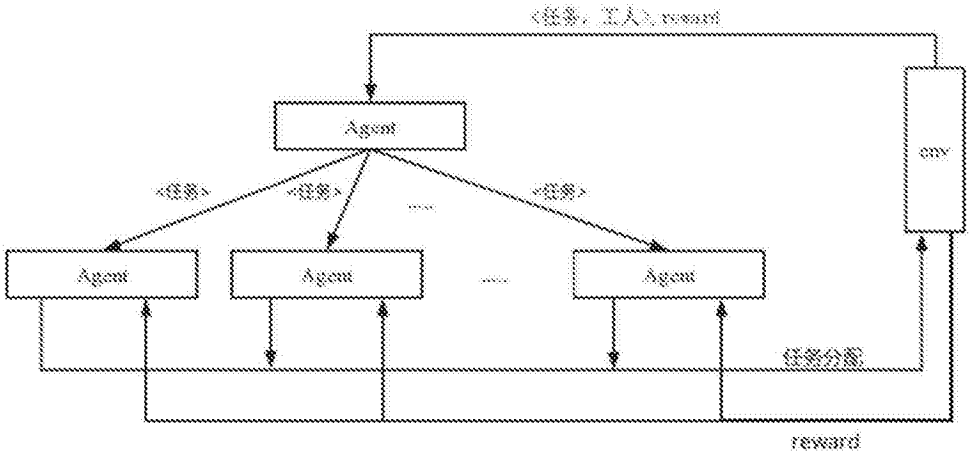

<任务，工人>,reward
Agent
env
<任务>
<任务>
<任务>
Agent
Agent
Agent
任务分配
reward

图2

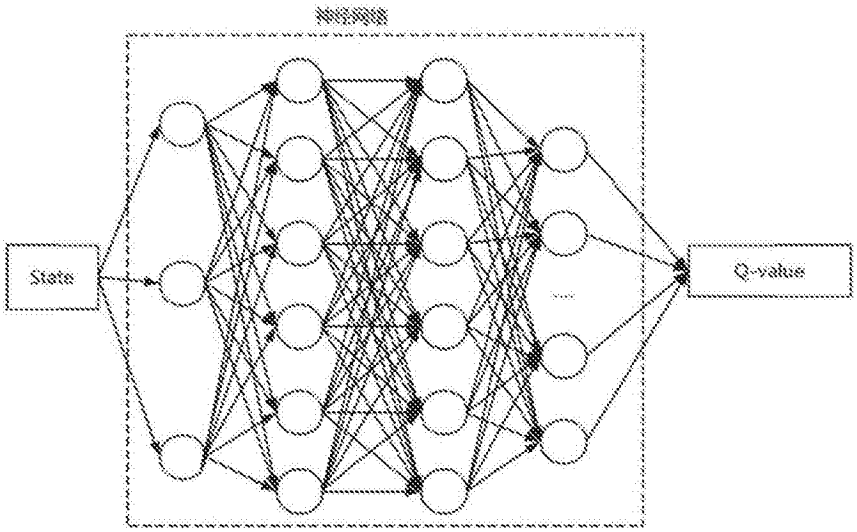

State
Q-value

图3