

缓解交通拥堵驾驶员控制系统、方法、介质、设备及应用

技术领域

本发明属于自动驾驶中的辅助驾驶技术领域，尤其涉及一种缓解交通拥堵驾驶员控制系统、方法、介质、设备及应用。

背景技术

目前，随着交通出行需求不断增长，城市道路网络的交通压力不断增加，交通拥堵问题日益严重。在城市路网中，拥堵的发生并不是独立的，当交通网络中某节点发生交通拥堵，如果不能及时疏散，会导致拥堵在整个网络中蔓延，继而引发大规模的网络拥堵，这种现象也被描述为相继拥堵。传统的解决思路是从增加基础设施建设入手，但是如果没有采取相应的规划与管理策略，拥堵疏导很难达到理想的效果。因此，分析拥堵在城市道路网络中的传播规律并制定合理、有效且适用性广的疏散策略，并且将最优疏散策略同时下发到当前区域内的每个车辆节点，对缓解交通拥堵，提高城市路网的交通疏导效率具有重要意义。

现有的涉及到交通拥堵的专利技术大多数停留在交通拥堵的感知、预测以及评估分析层面：

专利 CN113160570A 涉及一种交通拥堵预测方法及系统，通过获取待预测路网内所有车辆的历史轨迹数据，经过数据的处理以及整合得到所述待预测路网的邻接矩阵，将其输入到预测模型得到拥堵预测结果，能够提高交通拥堵的预测情况。

专利 CN112863176A 提供一种交通拥堵溯源方法，该方法对拥堵路径中各关键节点对拥堵的贡献度进行分析，以找出拥堵路径中的所有拥堵源，从而能够更精准和迅速对拥堵成因进行定位。

专利 CN113096397A 公开了一种基于毫米波雷达与视频检测的交通拥堵分

说明书

析方法，该发明利用视频监控摄像头获取视场内道路影像，并用毫米波雷达检测车辆位置和车辆行驶速度，根据以上交通流数据分析交通拥堵程度，实时、智能、快捷地对道路情况进行评估。

专利 CN113345238A 公开了一种基于城市公交网络感知的交通拥堵分析系统，该发明增加了拥堵验证模块，能够对拥有时间纵深的侧车辆以及同路段相关联车辆的数据进行验证，从而进一步提高拥堵预判分析的精准度。

专利 CN113256985A 公开了一种交通拥堵预测方法、装置和电子设备，对获取到的交通路况的道路特征数据进行两两车道间路况相似性处理，可以提高交通拥堵预测的精准度。

专利 CN113269768A 提供一种交通拥堵分析方法、装置及分析设备，属于图像处理技术领域，通过使用优化的深度学习网络模型进行目标检测，可以提高目标的检测精度和检测效率，从而使交通拥堵分析结果更加准确。

以上类型的现有专利在对于交通拥堵的感知、预测以及分析方面具有很好的效用，但对于交通拥堵问题始终没有实际展开进行解决，在现实交通系统中止步于问题的预测以及提出，因此其功用是不完善的。但是值得肯定的是，以上专利技术对于交通拥堵问题的实际解决具有重要意义，为系统的最优化解决方案的计算提供了稳定可靠的环境信息的输入。

近年来，也有相关的专利尝试对交通拥堵问题进行解决：

专利 CN113283714A 涉及一种基于群决策的交通拥堵抑制方法，属于交通拥堵抑制技术领域，利用网联环境中的终端传感器系统，收集并上传到云平台数据中心，利用城市各路口的交通监控系统采集路口交通数据，形成评价矩阵，并持续向云平台数据中心上传最新数据。在信息物理系统中道路拥堵抑制方面有明显效果，有效的减少道路拥堵阻塞，进而减少了行车的事故率；有效减少了交通行车中的停车率、等待时间，进而降低了出行成本。抑制了交通路网的拥堵，改善了城市干路的交通状况。

专利 CN112863204A 公开了一种缓解交通拥堵的智能城市通行优化系统，包括：通行数据采集机构、信号显示机构和支撑架，全面收集通行数据，灵活

说明书

改变交通信号，充分缓解交通拥堵的问题。

专利 CN111028505A 提供一种交通拥堵的治理方法及装置，涉及交通技术领域，能够提高交通拥堵的治理效率。目标地点为发生拥堵的地点；根据目标地点的历史交通流数据，确定目标地点的历史拥堵特征；根据目标地点的实时拥堵特征和历史拥堵特征，确定目标地点的拥堵治理优先级；根据目标区域的警力情况，以及多个目标地点中每一个目标地点的拥堵治理优先级，确定目标区域的警力分配方案。

专利 CN112348152A 公开一种基于改进层次分析法的交通拥堵疏导方法及系统，每次疏导都选取当前影响时间最短的备选路径推送给目标疏导车辆，直到所有目标疏导车辆疏导完成或者拥堵路段的预计影响时间小于当前备选路径的影响时间最小值。

以上类型的现有专利致力于通过预测交通拥堵的概率分布从而对其进行针对性的预防，以及通过对于交通流数据的实时检测对交通信号以及警力分布进行实时调整，从而辅助交通拥堵问题的解决。但是对于大型城市在交通高峰期必然出现的交通拥堵状况，以及在交通拥堵过程中常见的驾驶员行为的不确定性，将进一步提高系统的复杂度，以上专利方法均不适用。

综上，针对交通拥堵问题，在对其进行感知、预测以及评估的基础上，需要一套集成系统实时进行决策规划，给单车节点下发任务指令，从而实现全局利益的最优化，即以最高的效率以及安全性解决当前环境内的交通拥堵问题。

通过上述分析，现有技术存在的问题及缺陷为：

（1）传统的解决思路是从增加基础设施建设入手，但是如果没有采取相应的规划与管理策略，拥堵疏导很难达到理想的效果。

（2）现有技术对于交通拥堵问题始终没有实际展开进行解决，在现实交通系统中止步于问题的预测以及提出，因此其功用是不完善的。

（3）现有技术对于大型城市在交通高峰期必然出现的交通拥堵状况，以及在交通拥堵过程中常见的驾驶员行为的不确定性，将进一步提高系统的复杂度，以上专利方法均不适用。

说明书

解决以上问题及缺陷的难度为：在路口、岔路等存在大量交互过程的交通场景内，往往发生严重的交通拥堵问题，其根本原因在于分属于每个车辆节点的车辆驾驶员无法对于当前场景存在较为全局的认知，因此在自身利益效用最大化的驱使下，往往采取利己的驾驶策略，但是该过程往往会进一步加重交通拥堵问题；

在上述复杂博弈的交通场景内，每个单独车辆节点之间的相对关系错综复杂，其中包括相对位置关系、相对运动关系、竞争关系以及合作关系等，而在进行决策规划过程中，功能核心仍然为数学求解器，因此需要同时考量以上所有信息并进行量化，但是竞争合作等博弈关系具有动态不确定性，而且不同于传统的数据类型，因此难以进行有效量化。

解决以上问题及缺陷的意义为：整个方案采用集中运算、分别执行的模式，在进行场景内的全局规划决策过程中，综合考量场景内的全部车辆节点的信息，以全局最优化为目标，求解获得每个单车节点的最优动作选择，从而有效解决交通拥堵问题。

发明内容

针对现有技术存在的问题，本发明提供了一种缓解交通拥堵驾驶员控制系统、方法、介质、设备及应用，尤其涉及一种基于强化学习框架的缓解交通拥堵驾驶员控制系统、方法、介质、设备及应用，旨在缓解现有交通系统中路口存在的拥堵问题。

本发明是这样实现的，一种缓解交通拥堵驾驶员控制方法，所述缓解交通拥堵驾驶员控制方法包括：采用集中学习但分散执行的设定，每个目标车辆节点在每个时刻均做一个决策，为所有节点实现相同的给定目标，即有序通过从而解决交通拥堵问题；节点间的通信和信息传播采用图神经网络 GNN 进行建模，决策处理器采用 Deep Q learning，形成的决策信息以建议指令的形式下发给每一位环境内的驾驶员。

本发明采用集中学习、分散执行的设定有利于提高整个规划决策过程考量信息的全面性并将具体执行分配到单车节点由驾驶员完成最后一步任务，这样

说明书

可以提高计算效率的同时提高系统工作稳定性；为所有的节点实现相同给定目标是为了计算全局最优方案，避免由于单车节点采取局部最优行动进一步加重交通拥堵问题；单车节点之间的相对关系以及信息交互通过图神经网络进行信息处理，解决了场景内节点之间的博弈过程难以进行有效量化的问题；决策处理器采用深度强化学习，可以在缺少相关数据集的情况下依然可以通过与环境的交互完成训练任务，而且使用深度神经网络能够最大化地挖掘场景信息，从而使得决策过程更加全面可靠。

进一步，所述缓解交通拥堵驾驶员控制方法，还包括环境建模，对于环境的建模定义为信息拓扑结构；所述环境建模，包括：

根据车辆的空间位置和相对关系，将该模型分解为局部和全局两层，所述局部网络是一个“星形”图，包括目标车辆以及其周围的所有其他车辆个体；所述全局网络由当前环境内的所有车辆个体组成；目标车辆既通过车载传感器从附近的其他车辆获取局部信息，又通过云端连接通道从其他环境内的车辆获取全局信息。

进一步，在局部“星形”网络中，由于目标车辆具有传感功能，故信息从周围车辆传递到目标车辆；从全局网络来看，所有的目标车辆共享其他车辆的局部感知信息。

进一步，所述缓解交通拥堵驾驶员控制方法，还包括强化学习模型构建，所述强化学习模型结构，包括：

在每一个时间步 t ，在目标车辆周围的 N 个其他车辆均可被检测到，故每个时间步 t 对应的模型空间的输入设置为状态 s ，该状态 s 是一个由三个信息模块组成的元组，包括：节点特性 x_t 、关联矩阵 A_t ，记录车辆的索引矩阵 M_t ；其中节点特性 x_t 包括速度 v_t 、纵向位置 p_t 、横向车道位置 l_t 以及行驶意图 I_t ，关联矩阵 A_t 表示目标车辆及其周围车辆之间的交互关系，索引矩阵 M_t 用于将目标车辆从所有节点中筛选出来。

进一步，所述强化学习模型结构，还包括：

节点节点特性 x_t 输入到全连接 FCN 层，FCN 的输出与关联矩阵 A_t 同时输入

说明书

图神经网络 GCN 层进行并行计算,将输出结果与索引矩阵 M_i 进行点积运算进行车辆节点的筛选,最后通过 Q 网络计算输出 Q 值供参数进化迭代使用。

在正式训练开始前,设置 T 个时间步作为“热身阶段”;从 $T+1$ 时间步开始,按照奖励最大化以及损失最小化的原则对模型进行训练,用于保证模型训练的稳定性,使得车辆节点对环境进行探索。

本发明的另一目的在于提供一种应用所述的缓解交通拥堵驾驶员控制方法的缓解交通拥堵驾驶员控制系统,所述缓解交通拥堵驾驶员控制系统,包括:

环境模型构建模块,用于根据车辆的空间位置和相对关系,将环境模型分解为局部和全局两层,并将环境的建模定义为信息拓扑结构;

强化学习模型构建模块,用于将节点节点特性 x_i 输入到全连接 FCN 层,FCN 的输出与关联矩阵 A_i 同时输入图神经网络 GCN 层进行并行计算,将输出结果与索引矩阵 M_i 进行点积运算进行车辆节点的筛选,最后通过 Q 网络计算输出 Q 值供参数进化迭代使用;

交通拥堵缓解模块,用于采用集中学习但分散执行的设定,每个目标车辆节点在每个时刻均做一个决策,为所有节点实现相同的给定目标,即有序通过从而解决交通拥堵问题;

决策信息建立下发模块,用于将节点间的通信和信息传播采用图神经网络 GNN 进行建模,决策处理器采用 Deep Q learning,形成的决策信息以建议指令的形式下发给每一位环境内的驾驶员。

本发明的另一目的在于提供一种计算机设备,所述计算机设备包括存储器和处理器,所述存储器存储有计算机程序,所述计算机程序被所述处理器执行时,使得所述处理器执行如下步骤:

采用集中学习但分散执行的设定,每个目标车辆节点在每个时刻均做一个决策,为所有节点实现相同的给定目标,即有序通过从而解决交通拥堵问题;节点间的通信和信息传播采用图神经网络 GNN 进行建模,决策处理器采用 Deep Q learning,形成的决策信息以建议指令的形式下发给每一位环境内的驾驶员。

本发明的另一目的在于提供一种计算机可读存储介质,存储有计算机程序,

说明书

所述计算机程序被处理器执行时，使得所述处理器执行如下步骤：

采用集中学习但分散执行的设定，每个目标车辆节点在每个时刻均做一个决策，为所有节点实现相同的给定目标，即有序通过从而解决交通拥堵问题；节点间的通信和信息传播采用图神经网络 GNN 进行建模，决策处理器采用 Deep Q learning，形成的决策信息以建议指令的形式下发给每一位环境内的驾驶员。

本发明的另一目的在于提供一种信息数据处理终端，所述信息数据处理终端用于实现所述的缓解交通拥堵驾驶员控制系统。

本发明的另一目的在于提供一种应用所述的缓解交通拥堵驾驶员控制方法的缓解交通拥堵驾驶员控制方法的应用方法，所述应用方法，包括：

将感知系统获得的局部信息以及全局信息进行初步处理可以获得符合网络输入要求的数据类型，将此数据元组输入到训练成熟的网络获得全局的最优决策输出，再将决策结果通过驾驶员建议系统下发到每个车辆节点的实际操作人，由实际操作人完成最后的车辆行驶控制任务，从而有效解决交通拥堵问题。

结合上述的所有技术方案，本发明所具备的优点及积极效果为：本发明提供的缓解交通拥堵驾驶员控制方法，基于强化学习框架，解决了现有技术对于已经发生的交通拥堵无法进行直接性以及根本性解决的问题。对于已经发生或即将发生交通拥堵的场景做出实时的最优化决策规划方案，使用强化学习以及图神经网络结合的方式进行规划决策，并用于处理交通拥堵问题是本发明最先提出，具有很强的创新意义。该方案细化到当前环境内所有单车节点的下一步行动，与目前已有的方法相比，本发明的有益效果可以具化为以下几个方面：

1. 本发明提出的基于强化学习算法的缓解交通拥堵的方法具备场景遍历深度的优势，针对某一细分场景，通过大数据系统更容易覆盖全部工况；本发明提出的基于强化学习算法的缓解交通拥堵的方法具备自学习性能，机器能够自行提炼环境特征和决策属性，便于系统优化迭代；本发明提出的基于强化学习算法的缓解交通拥堵的方法不必遍历各种工况，通过数据的训练完善模型，模型正确可率随数据的完备得以提升。

2. 基于规则的算法面临复杂工况，不可避免因遍历工况、完善逻辑而导致

说明书

算法规模庞大冗杂，与之相比，在系统复杂性上，本发明提出的基于强化学习算法的缓解交通拥堵的方法舍弃了基于规则算法的层级架构，更加简洁直接，大幅简化了决策系统结构，更适用于复杂工况。

3. 本发明将当前场景内的所有单车节点之间的关系转化为拓扑图信息作为基本信息外的信息补充用于系统的输入，使得系统具有全局最优化的能力；本发明将决策系统的输出结果通过对驾驶员建议的方式进行任务发布，从而使得最底层的规划问题由驾驶员来完成，这样可以大大提高整个系统的运行安全性以及稳定性，使单车节点同时具有全局规划最优以及局部规划最优的特性。

4. 本发明在进行问题解决（即交通拥堵的疏通）的过程中，跳过了单个节点的收益，采用全局最优的方案进行决策的制定，这能有效针对交通拥堵这一系统性的问题进行解决；在进行数据处理的过程中重复考虑了环境内单个车辆节点与环境之间，以及节点与节点之间的相互作用关系，对于全局最优解的制定具有重要意义；使用强化学习训练获得的网络模型作为系统的求解器，可以借助大数据系统从而更好地覆盖全部工况。

5. 本发明在模型训练的初期设置热“热身阶段”，通过不断试错的方式让模型得以充分探索环境，从而提高其稳定性以及应用安全性；将获得的全局最优决策结果通过驾驶员建议系统下发给车辆节点的实际操作人，最后的车辆操作控制由真人来完成，实现了车联网系统全局感知广度特性与真人实际操作稳定性的互补。

附图说明

为了更清楚地说明本发明实施例的技术方案，下面将对本发明实施例中所需要使用的附图做简单的介绍，显而易见地，下面所描述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。

图 1 是本发明实施例提供的缓解交通拥堵驾驶员控制方法流程图。

图 2 是本发明实施例提供的缓解交通拥堵驾驶员控制系统结构框图；

说明书

图中：1、环境模型构建模块；2、强化学习模型构建模块；3、交通拥堵缓解模块；4、决策信息建立下发模块。

图 3 是本发明实施例提供的马尔可夫博弈过程的示意图。

图 4 是本发明实施例提供的对于实际应用场景的环境建模示意图。

图 5 是本发明实施例提供的基于强化学习算法的缓解交通拥堵的方法中所使用的强化学习模型的示意图。

图 6 是本发明实施例提供的在实际交通拥堵环境中应用的流程图。

具体实施方式

为了使本发明的目的、技术方案及优点更加清楚明白，以下结合实施例，对本发明进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

针对现有技术存在的问题，本发明提供了一种缓解交通拥堵驾驶员控制系统、方法、介质、设备及应用，下面结合附图对本发明作详细的描述。

如图 1 所示，本发明实施例提供的缓解交通拥堵驾驶员控制方法包括以下步骤：

S101，分别进行环境模型以及强化学习模型的构建；

S102，采用集中学习但分散执行的设定，每个目标车辆节点在每个时刻均做一个决策，为所有节点实现相同的给定目标，即有序通过解决交通拥堵问题；

S103，节点间的通信和信息传播采用图神经网络GNN进行建模，决策处理器采用Deep Q learning，形成的决策信息以建议指令的形式下发给每一位环境内的驾驶员。

如图 2 所示，本发明实施例提供的缓解交通拥堵驾驶员控制系统，包括：

环境模型构建模块 1，用于根据车辆的空间位置和相对关系，将环境模型分解为局部和全局两层，并将环境的建模定义为信息拓扑结构；

强化学习模型构建模块 2，用于将节点节点特性 x_i 输入到全连接 FCN 层，FCN 的输出与关联矩阵 A_i 同时输入图神经网络 GCN 层进行并行计算，将输出结

说明书

果与索引矩阵 M_i 进行点积运算进行车辆节点的筛选，最后通过Q网络计算输出Q值供参数进化迭代使用；

交通拥堵缓解模块3，用于采用集中学习但分散执行的设定，每个目标车辆节点在每个时刻均做一个决策，为所有节点实现相同的给定目标，即有序通过从而解决交通拥堵问题；

决策信息建立下发模块4，用于将节点间的通信和信息传播采用图神经网络GNN进行建模，决策处理器采用Deep Q learning，形成的决策信息以建议指令的形式下发给每一位环境内的驾驶员。

下面结合具体实施例对本发明的技术方案作进一步描述。

本发明提供了一种基于强化学习框架缓解交通拥堵驾驶员控制系统，解决现有技术对于已经发生的交通拥堵无法进行直接性以及根本性解决的问题。

由于本发明针对的交通拥堵场景内移动同时存在多个车辆个体，多个车辆节点不仅会与环境进行交互，而且节点之间也会存在较强的相互作用，因此整个系统可以归纳为一个马尔可夫博弈，如图3所示。

本发明提出的基于强化学习算法的缓解交通拥堵的方法对于环境的建模如图4所示，根据车辆的空间位置和相对关系，将该模型进一步分解为局部和全局两层：局部网络是一个“星形”图，包括目标车辆以及其周围的所有其他车辆个体；而全局网络由当前环境内的所有车辆个体组成。目标车辆既通过车载传感器从附近的其他车辆获取局部信息，又通过云端连接通道从其他环境内的车辆获取全局信息。本方法内对于环境的建模可以定义为信息拓扑结构。

在局部“星形”网络中，信息从周围车辆传递到目标车辆，因为目标车辆具有传感功能。从全局网络来看，所有的目标车辆可以共享其他车辆的局部感知信息。

本发明提出的基于强化学习算法的缓解交通拥堵的方法采用集中学习但分散执行的设定。在这种设定中，每个目标车辆节点在每个时刻都要做一个决策，目标是为所有节点实现相同的给定目标--即有序通过从而解决交通拥堵问题。节点间的通信和信息传播采用图神经网络(GNN)进行建模，决策处理器采用Deep

说明书

Q learning, 最后形成的决策信息以建议指令的形式下发给每一位环境内驾驶员。

强化学习模型结构

在每一个时间步 t , 在目标车辆周围的 N 个其他车辆均可以被检测到, 因此每个时间步 t 对应的模型空间的输入设置为状态 s , 该状态 s 是一个由三个信息模块组成的元组, 包括: 节点特性 x_t 、关联矩阵 A_t , 记录车辆的索引矩阵 M_t , 其中节点特性 x_t 包括速度 v_t 、纵向位置 p_t 、横向车道位置 l_t 以及行驶意图 l_t , 关联矩阵 A_t 表示目标车辆及其周围车辆之间的交互关系, 索引矩阵 M_t 用于将目标车辆从所有节点中筛选出来。

整体的模型结构如图 5 所示, 首先节点节点特性 x_t 输入到全连接 (FCN) 层, FCN 的输出与关联矩阵 A_t 同时输入图神经网络 (GCN) 层进行并行计算, 将输出结果与索引矩阵 M_t 进行点积运算进行车辆节点的筛选, 最后通过 Q 网络计算输出 Q 值供参数进化迭代使用。

为了保证模型训练的稳定性, 与此同时使得车辆节点能够充分地对环境进行探索, 在正式训练开始之前设置了 r 个时间步作为“热身阶段”, 这有助于系统保证决策的安全性。从 $r+1$ 时间步开始, 按照奖励最大化以及损失最小化的原则对模型进行训练。

本发明提出的基于强化学习算法的缓解交通拥堵的方法的实际应用流程如图 6 所示, 将感知系统获得的局部信息以及全局信息进行初步处理可以获得符合网络输入要求的数据类型, 将此数据元组输入到训练成熟的网络获得全局的最优决策输出, 再将决策结果通过驾驶员建议系统下发到每个车辆节点的实际操作人, 由实际操作人完成最后的车辆行驶控制任务, 从而有效解决交通拥堵问题。

本发明实施例包括在数据采集的过程结束后, 利用拓扑图的形式对环境内车辆节点之间的相对关系进行建模, 从而形成关联矩阵对节点间的博弈关系进行量化的过程。

本发明实施例包括但不限于使用 DQN 进行后续模型学习过程的实现, 还可以使用卷积神经网络 (CNN)、深度置信网络 (DBN)、受限玻尔兹曼机 (RBM)、

说明书

递归神经网络 (RNN & LSTM & GRU)、递归张量神经网络 (RNTN)、自动编码器 (AutoEncoder)、生成对抗网络 (GAN)等等形式的网络进行功能的实现。

本发明实施例在获取全局效用最优解后，形成的单个节点的最优驾驶行为，包括但不限于通过驾驶员建议系统进行指令下达，还可以通过集成在手机软件内对车辆驾驶员传达最优指令。对于具有 L2 及以上自动驾驶能力的汽车，可以通过驾驶员辅助系统以方向盘以及脚踏板力反馈的方式进行指令啊下达以及传递。对于完全自主驾驶的无人车辆平台，可直接通过车辆节点与终端之间的信息传递直接将指令解算为车辆的纵向加速度以及横摆角加速度实现预期动作；

为了证明本发明提出的一种基于强化学习框架缓解交通拥堵的驾驶员建议系统的规划决策能力，可以借助 SUMO 仿真平台对高速公路岔路口进行环境建模，使用现有的基于规则的规划决策方法 (Rule-Based) 与本发明提出的方法 (简称为 GCQ) 进行全局奖励的比较，结果如表 1 所示。

表 1

HDV Inflow Model		0.1 veh/sec	0.2 veh/sec	0.3 veh/sec	0.4 veh/sec	0.5 veh/sec
GCQ	mean	3521.86	3537.71	3702.59	2594.94	4152.41
	median	3509.65	3635.83	3728.61	3288.57	4153.65
	std	491.77	1121.65	627.58	1815.32	447.03
Rule-based	mean	-10040.15	-6953.08	-6762.59	-9221.64	-8192.19
	median	-9327.27	-6281.48	-6227.01	-8368.58	-7811.41
	std	5788.20	4026.39	3678.88	4538.56	3412.66

可以发现，随着环境内车辆节点数量 (veh/sec) 的增多，本发明提出的 GCQ 算法获得的全局奖励值在平均值 (mean)、中间值 (median)、标准差 (std) 等衡量指标上均大大优于基于规则的算法。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用全部或部分地以计算机程序产品的形式实现，所述计算机程序产品包括一个或多个计算机指令。在计算机上加载或执行所述计算机程序指令时，全部或部分地产生按照本发明实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线 (例如同轴电缆、光纤、数字用户线 (DSL))

说明书

或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务器或数据中心进行传输）。所述计算机可读取存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质（例如软盘、硬盘、磁带）、光介质（例如 DVD）、或者半导体介质（例如固态硬盘 Solid State Disk（SSD））等。

以上所述，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等，都应涵盖在本发明的保护范围之内。