



(12) 发明专利申请

(10) 申请公布号 CN 117649168 A

(43) 申请公布日 2024. 03. 05

(21) 申请号 202210966505.8

G06N 3/08 (2023.01)

(22) 申请日 2022.08.12

(71) 申请人 株式会社日立制作所

地址 日本东京都

申请人 清华大学

(72) 发明人 李勇 宗泽方 颜欢 郑萌

庄云麟 耿璐

(74) 专利代理机构 北京银龙知识产权代理有限公司

公司 11243

专利代理师 姜精斌

(51) Int. Cl.

G06Q 10/0835 (2023.01)

G06F 18/25 (2023.01)

G06N 3/0455 (2023.01)

G06N 3/0464 (2023.01)

权利要求书4页 说明书18页 附图7页

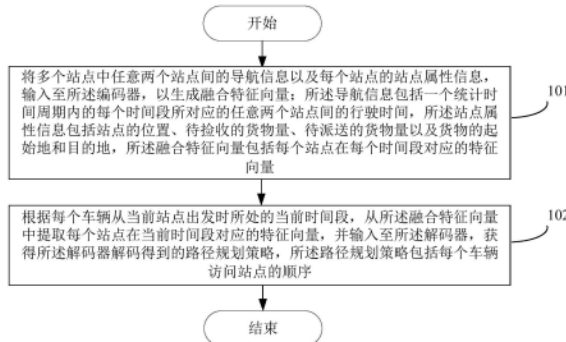
(54) 发明名称

一种时间依赖的路径规划方法及装置

(57) 摘要

本申请提供一种时间依赖的路径规划方法及装置,属于车辆路径规划技术领域。方法包括:将多个站点中任意两个站点间的导航信息以及每个站点的站点属性信息,输入至所述编码器,以生成融合特征向量;所述导航信息包括一个统计时间周期内的每个时间段所对应的任意两个站点间的行驶时间,所述站点属性信息包括站点的位置、待接收的货物量、待派送的货物量以及货物的起始地和目的地,所述融合特征向量包括每个站点在每个时间段对应的特征向量;

根据每个车辆从当前站点出发时所处的当前时间段,从所述融合特征向量中提取每个站点在当前时间段对应的特征向量,并输入至所述解码器,获得所述解码器解码得到的路径规划策略,所述路径规划策略包括每个车辆访问站点的顺序。本申请考虑了与时间相关的站点间行驶时间,能够获得更符合实际交通情况的路径规划策略,并能提高路径规划的效率。



1. 一种时间依赖的路径规划方法, 其特征在于, 用于利用预先训练的策略模型生成多个车辆访问多个站点的路径, 所述策略模型包括有编码器和解码器, 所述多个站点包括待派送站点和待接收站点, 所述方法包括:

编码步骤, 将多个站点中任意两个站点间的导航信息以及每个站点的站点属性信息, 输入至所述编码器, 以生成融合特征向量; 其中, 所述导航信息包括一个统计时间周期内的每个时间段所对应的任意两个站点间的行驶时间, 所述站点属性信息包括站点的位置、待接收的货物量、待派送的货物量以及货物的起始地和目的地, 所述融合特征向量包括每个站点在每个时间段对应的特征向量;

解码步骤, 根据每个车辆从当前站点出发时所处的当前时间段, 从所述融合特征向量中提取每个站点在当前时间段对应的特征向量, 并输入至所述解码器, 获得所述解码器解码得到的路径规划策略, 所述路径规划策略包括每个车辆访问站点的顺序。

2. 根据权利要求1所述的路径规划方法, 其特征在于,

所述编码器包括第一网络、第二网络和融合模块, 其中:

所述第一网络根据所述多个站点中任意两个站点间的导航信息, 生成每个站点的第一特征向量;

所述第二网络根据所述多个站点中每个站点的站点属性信息, 生成每个站点的第二特征向量;

所述融合模块融合每个站点的第一特征向量和第二特征向量, 获得所述融合特征向量。

3. 根据权利要求2所述的路径规划方法, 其特征在于,

所述第二网络还根据所述多个站点中每个站点的站点属性信息以及任意两个站点间的导航信息, 生成每个站点的第二特征向量。

4. 根据权利要求3所述的路径规划方法, 其特征在于,

所述第一网络为采用了动态池化的全卷积网络, 所述第二网络为基于注意力机制的神经网络。

5. 根据权利要求1所述的路径规划方法, 其特征在于,

所述解码器包括交互模块、与车辆一一对应的解码模块、矛盾处理模块和策略输出模块; 在每轮车辆调度中:

每个车辆对应的解码模块, 根据该车辆从当前站点出发时所处的当前时间段, 从所述融合特征向量中提取每个站点在当前时间段对应的特征向量, 根据所提取的特征向量和该车辆所在的当前站点, 计算该车辆去往各个站点的概率, 确定该车辆去往的下一站点;

所述矛盾处理模块在所有解码模块都确定了对应车辆去往的下一站点后, 若存在至少两个车辆去往同一个下一站点, 则选择所述至少两个车辆中的一个车辆去往该下一站点, 其他车辆停留在原站点;

所述交互模块根据各个车辆去往的下一站点, 更新每个车辆访问的站点列表、各个站点的访问状态、各个车辆所在的当前站点和剩余容量、以及各个车辆从当前站点出发时所处的当前时间段;

所述策略输出模块根据更新后的各个站点的访问状态, 判断所有站点是否都已经被访问, 若是, 则将全部车辆调回始发站点, 并输出所述路径规划策略, 否则, 继续由每个车辆对

应的解码模块进行下一轮车辆调度。

6. 根据权利要求5所述的路径规划方法,其特征在于,

所述待派送站点与所述待检收站点一一对应;

所述确定该车辆去往的下一站点,包括:

将访问状态为已访问的站点设为无法访问的站点;

将该车辆未访问的待检收站点所对应的待派送站点设为无法访问的站点;

将待检收的货物量超过该车辆的剩余容量的站点设为无法访问的站点;

删除无法访问的站点,将剩余站点中概率最大的站点,确定为该车辆去往的下一站点。

7. 根据权利要求5所述的路径规划方法,其特征在于,

所述解码模块为基于注意力机制的神经网络。

8. 根据权利要求1至7任一项所述的路径规划方法,其特征在于,还包括训练所述策略模型的步骤,训练所述策略模型的步骤包括:

获取步骤,获取多组训练数据,每组所述训练数据包括每个站点的位置、所述多个站点中任意两个站点间的导航信息、每个站点待投递的货物数量、每个站点待拾取的货物数量、货物的起始地和目的地、所述多个车辆中各个车辆的容量以及车辆的数目;

初始化步骤,初始化策略模型参数;

训练步骤,搭建策略模型,将所述多组训练数据输入所述策略模型,输出与每组训练数据对应的派送策略序列以及每个派送策略中各客户站点被选取的概率;利用所述派送策略序列计算所有车辆的总成本,以及局部解对应的局部成本;计算所述派送策略序列的评估值函数网络模型,所述评估值函数网络模型的输入为所述派送策略序列的每步局部解,输出为每步局部评估值;基于值函数更新所述策略模型和所述评估值函数;

重复所述训练步骤,直至满足预设的训练结束条件后,保存策略模型。

9. 根据权利要求8所述的路径规划方法,其特征在于,所述基于值函数更新所述策略模型和所述评估值函数包括:

计算所述策略模型和所述评估值函数的损失函数 L ;

对所述损失函数 L 求梯度;

利用梯度下降法更新所述策略模型和所述评估值函数的参数。

10. 一种时间依赖的路径规划装置,其特征在于,用于利用预先训练的策略模型生成多个车辆访问多个站点的路径,所述策略模型包括有编码器和解码器,所述多个站点包括待派送站点和待检收站点,所述装置包括:

信息输入单元,用于将多个站点中任意两个站点间的导航信息以及每个站点的站点属性信息,输入至所述编码器,以生成融合特征向量;其中,所述导航信息包括一个统计时间周期内的每个时间段所对应的任意两个站点间的行驶时间,所述站点属性信息包括站点的位置、待检收的货物量、待派送的货物量以及货物的起始地和目的地,所述融合特征向量包括每个站点在每个时间段对应的特征向量;

策略输出单元,用于根据每个车辆从当前站点出发时所处的当前时间段,从所述融合特征向量中提取每个站点在当前时间段对应的特征向量,并输入至所述解码器,获得所述解码器解码得到的路径规划策略,所述路径规划策略包括每个车辆访问站点的顺序。

11. 根据权利要求10所述的路径规划装置,其特征在于,

所述编码器包括第一网络、第二网络和融合模块,其中:

所述第一网络,用于根据所述多个站点中任意两个站点间的导航信息,生成每个站点的第一特征向量;

所述第二网络,用于根据所述多个站点中每个站点的站点属性信息,生成每个站点的第二特征向量;

所述融合模块,用于融合每个站点的第一特征向量和第二特征向量,获得所述融合特征向量。

12. 根据权利要求11所述的路径规划装置,其特征在于,

所述第一网络为采用了动态池化的全卷积网络,所述第二网络为基于注意力机制的神经网络。

13. 根据权利要求10所述的路径规划装置,其特征在于,

所述解码器包括交互模块、与车辆一一对应的解码模块、矛盾处理模块和策略输出模块;

每个车辆对应的解码模块,用于在每轮车辆调度中,根据该车辆从当前站点出发时所处的当前时间段,从所述融合特征向量中提取每个站点在当前时间段对应的特征向量,根据所提取的特征向量和该车辆所在的当前站点,计算该车辆去往各个站点的概率,确定该车辆去往的下一站点;

所述矛盾处理模块,用于在每轮车辆调度中,在所有解码模块都确定了对应车辆去往的下一站点后,若存在至少两个车辆去往同一个下一站点,则选择所述至少两个车辆中的一个车辆去往该下一站点,其他车辆停留在原站点;

所述交互模块,用于在每轮车辆调度中,根据各个车辆去往的下一站点,更新每个车辆访问的站点列表、各个站点的访问状态、各个车辆所在的当前站点和剩余容量、以及各个车辆从当前站点出发时所处的当前时间段;

所述策略输出模块,用于根据更新后的各个站点的访问状态,判断所有站点是否都被访问,若是,则将全部车辆调回始发站点,并输出所述路径规划策略,否则,继续由每个车辆对应的解码模块进行下一轮车辆调度。

14. 根据权利要求13所述的路径规划装置,其特征在于,

所述解码模块为基于注意力机制的神经网络。

15. 根据权利要求10至14任一项所述的路径规划装置,其特征在于,还包括训练单元,用于训练所述策略模型,所述训练单元包括:

获取子单元,用于获取多组训练数据,每组所述训练数据包括每个站点的位置、所述多个站点中任意两个站点间的导航信息、每个站点待投递的货物数量、每个站点待拾取的货物数量、货物的起始地和目的地、所述多个车辆中各个车辆的容量以及车辆的数目;

初始化子单元,初始化策略模型的参数;

训练子单元,用于搭建策略模型,将所述多组训练数据输入所述策略模型,输出与每组训练数据对应的派送策略序列以及每个派送策略中各客户站点被选取的概率;利用所述派送策略序列计算所有车辆的总成本,以及局部解对应的局部成本;计算所述派送策略序列的评估值函数网络模型,所述评估值函数网络模型的输入为所述派送策略序列的每步局部解,输出为每步局部评估值;基于值函数更新所述策略模型和所述评估值函数;

保存子单元,用于重复所述训练步骤,直至满足预设的训练结束条件后,保存策略模型。

一种时间依赖的路径规划方法及装置

技术领域

[0001] 本申请涉及物流配送中的车辆路径规划技术领域,特别是指一种时间依赖的路径规划方法及装置。

背景技术

[0002] 物流行业是一种兼顾成本和服务质量的服务行业。在为客户提供物流配送服务时,物流服务提供者需要提供高效的准时的服务,同时尽可能的降低服务的成本。

[0003] 物流配送是物流服务的关键环节,是指根据客户对货物运输的时间和运量等的要求,将货物从出发地运送到目的地的过程。物流服务提供者一般在满足客户要求的前提下,为了降低配送成本,尽量减少配送用车辆的行驶时间。

[0004] 车辆路径规划问题是解决物流配送的经典问题,也是运筹学的经典问题,其目标是给定一组具有不同配送需求量的站点和一组有容量等限制的配送车辆,求取完成配送时配送车辆的行驶路径,使所有车辆总的行驶时间(行驶路径)最短。

[0005] 现有技术的一种车辆路径规划方法,通过启发式算法来进行多车辆路径规划,但派送策略的生成速度较慢。现有技术的另一种时间依赖的路径规划方法,是采用强化学习训练一个解决多车辆路径规划的模型,并在实际应用时通过模型的推导能力直接给出输入问题的派送策略。但这样的模型通常以站点间的欧式距离作为度量,无法考虑车辆在不同站点之间按照路网行驶路线的距离,也没有建模时间依赖的行驶距离和时间的能力,因此所得到的路线可能并非是成本最优的。这严重限制了模型的应用场景。现有技术的另一种时间依赖的路径规划方法,其构造了一个全连接的有向图来模拟VRP问题,在此有向图上使用启发式算法搜索最优解,存在计算量大效率低等问题。这里,时间依赖是指在所有站点中,任意两个站点之间的行驶距离和时间可能随时间段的不同而改变,进一步提升了问题的复杂度。

发明内容

[0006] 本申请要解决的技术问题是提供一种时间依赖的路径规划方法及装置,考虑了与时间相关的站点间行驶时间,使得生成的路径规划策略更符合实际交通情况,并能够提高提高路径规划的效率。

[0007] 为解决上述技术问题,本申请的实施例提供技术方案如下:

[0008] 一方面,提供一种时间依赖的路径规划方法,其特征在于,用于利用预先训练的策略模型生成多个车辆访问多个站点的路径,所述策略模型包括有编码器和解码器,所述多个站点包括待派送站点和待接收站点,所述方法包括:

[0009] 编码步骤,将多个站点中任意两个站点间的导航信息以及每个站点的站点属性信息,输入至所述编码器,以生成融合特征向量;其中,所述导航信息包括一个统计时间周期内的每个时间段所对应的任意两个站点间的行驶时间,所述站点属性信息包括站点的位置、待接收的货物量、待派送的货物量以及货物的起始地和目的地,所述融合特征向量包括

每个站点在每个时间段对应的特征向量；

[0010] 解码步骤,根据每个车辆从当前站点出发时所处的当前时间段,从所述融合特征向量中提取每个站点在当前时间段对应的特征向量,并输入至所述解码器,获得所述解码器解码得到的路径规划策略,所述路径规划策略包括每个车辆访问站点的顺序。

[0011] 本申请的可选实施例中,所述编码器包括第一网络、第二网络和融合模块,其中:

[0012] 所述第一网络根据所述多个站点中任意两个站点间的导航信息,生成每个站点的第一特征向量;

[0013] 所述第二网络根据所述多个站点中每个站点的站点属性信息,生成每个站点的第二特征向量;

[0014] 所述融合模块融合每个站点的第一特征向量和第二特征向量,获得所述融合特征向量。

[0015] 本申请的可选实施例中,所述第二网络还根据所述多个站点中每个站点的站点属性信息以及任意两个站点间的导航信息,生成每个站点的第二特征向量。

[0016] 本申请的可选实施例中,所述第一网络为采用了动态池化的全卷积网络,所述第二网络为基于注意力机制的神经网络。

[0017] 本申请的可选实施例中,所述解码器包括交互模块、与车辆一一对应的解码模块、矛盾处理模块和策略输出模块;在每轮车辆调度中:

[0018] 每个车辆对应的解码模块,根据该车辆从当前站点出发时所处的当前时间段,从所述融合特征向量中提取每个站点在当前时间段对应的特征向量,根据所提取的特征向量和该车辆所在的当前站点,计算该车辆去往各个站点的概率,确定该车辆去往的下一站点;

[0019] 所述矛盾处理模块在所有解码模块都确定了对应车辆去往的下一站点后,若存在至少两个车辆去往同一个下一站点,则选择所述至少两个车辆中的一个车辆去往该下一站点,其他车辆停留在原站点;

[0020] 所述交互模块根据各个车辆去往的下一站点,更新每个车辆访问的站点列表、各个站点的访问状态、各个车辆所在的当前站点和剩余容量、以及各个车辆从当前站点出发时所处的当前时间段;

[0021] 所述策略输出模块根据更新后的各个站点的访问状态,判断所有站点是否都被访问,若是,则将全部车辆调回始发站点,并输出所述路径规划策略,否则,继续由每个车辆对应的解码模块进行下一轮车辆调度。

[0022] 本申请的可选实施例中,所述待派送站点与所述待接收站点一一对应;

[0023] 所述确定该车辆去往的下一站点,包括:

[0024] 将访问状态为已访问的站点设为无法访问的站点;

[0025] 将该车辆未访问的待接收站点所对应的待派送站点设为无法访问的站点;

[0026] 将待接收的货物量超过该车辆的剩余容量的站点设为无法访问的站点;

[0027] 删除无法访问的站点,将剩余站点中概率最大的站点,确定为该车辆去往的下一站点。

[0028] 本申请的可选实施例中,所述解码模块为基于注意力机制的神经网络。

[0029] 本申请的可选实施例中,还包括训练所述策略模型的步骤,训练所述策略模型的步骤包括:

[0030] 获取步骤,获取多组训练数据,每组所述训练数据包括每个站点的位置、所述多个站点中任意两个站点间的导航信息、每个站点待投递的货物数量、每个站点待拾取的货物数量、货物的起始地和目的地、所述多个车辆中各个车辆的容量以及车辆的数目;

[0031] 初始化步骤,初始化策略模型的参数;

[0032] 训练步骤,搭建策略模型,将所述多组训练数据输入所述策略模型,输出与每组训练数据对应的派送策略序列以及每个派送策略中各客户站点被选取的概率;利用所述派送策略序列计算所有车辆的总成本,以及局部解对应的局部成本;计算所述派送策略序列的评估值函数网络模型,所述评估值函数网络模型的输入为所述派送策略序列的每步局部解,输出为每步局部评估值;基于值函数更新所述策略模型和所述评估值函数;

[0033] 重复所述训练步骤,直至满足预设的训练结束条件后,保存策略模型。

[0034] 本申请的可选实施例中,所述基于值函数更新所述策略模型和所述评估值函数包括:

[0035] 计算所述策略模型和所述评估值函数的损失函数L;

[0036] 对所述损失函数L求梯度;

[0037] 利用梯度下降法更新所述策略模型和所述评估值函数的参数。

[0038] 另一方面,本申请实施例还提供了一种时间依赖的路径规划装置,用于利用预先训练的策略模型生成多个车辆访问多个站点的路径,所述策略模型包括有编码器和解码器,所述多个站点包括待派送站点和待接收站点,所述装置包括:

[0039] 信息输入单元,用于将多个站点中任意两个站点间的导航信息以及每个站点的站点属性信息,输入至所述编码器,以生成融合特征向量;其中,所述导航信息包括一个统计时间周期内的每个时间段所对应的任意两个站点间的行驶时间,所述站点属性信息包括站点的位置、待接收的货物量、待派送的货物量以及货物的起始地和目的地,所述融合特征向量包括每个站点在每个时间段对应的特征向量;

[0040] 策略输出单元,用于根据每个车辆从当前站点出发时所处的当前时间段,从所述融合特征向量中提取每个站点在当前时间段对应的特征向量,并输入至所述解码器,获得所述解码器解码得到的路径规划策略,所述路径规划策略包括每个车辆访问站点的顺序。

[0041] 本申请的可选实施例中,所述编码器包括第一网络、第二网络和融合模块,其中:

[0042] 所述第一网络,用于根据所述多个站点中任意两个站点间的导航信息,生成每个站点的第一特征向量;

[0043] 所述第二网络,用于根据所述多个站点中每个站点的站点属性信息,生成每个站点的第二特征向量;

[0044] 所述融合模块,用于融合每个站点的第一特征向量和第二特征向量,获得所述融合特征向量。

[0045] 本申请的可选实施例中,所述第二网络,还用于根据所述多个站点中每个站点的站点属性信息以及任意两个站点间的导航信息,生成每个站点的第二特征向量。

[0046] 本申请的可选实施例中,所述第一网络为采用了动态池化的全卷积网络,所述第二网络为基于注意力机制的神经网络。

[0047] 本申请的可选实施例中,所述解码器包括交互模块、与车辆一一对应的解码模块、矛盾处理模块和策略输出模块;

[0048] 每个车辆对应的解码模块,用于在每轮车辆调度中,根据该车辆从当前站点出发时所处的当前时间段,从所述融合特征向量中提取每个站点在当前时间段对应的特征向量,根据所提取的特征向量和该车辆所在的当前站点,计算该车辆去往各个站点的概率,确定该车辆去往的下一站点;

[0049] 所述矛盾处理模块,用于在每轮车辆调度中,在所有解码模块都确定了对应车辆去往的下一站点后,若存在至少两个车辆去往同一个下一站点,则选择所述至少两个车辆中的一个车辆去往该下一站点,其他车辆停留在原站点;

[0050] 所述交互模块,用于在每轮车辆调度中,根据各个车辆去往的下一站点,更新每个车辆访问的站点列表、各个站点的访问状态、各个车辆所在的当前站点和剩余容量、以及各个车辆从当前站点出发时所处的当前时间段;

[0051] 所述策略输出模块,用于根据更新后的各个站点的访问状态,判断所有站点是否都被访问,若是,则将全部车辆调回始发站点,并输出所述路径规划策略,否则,继续由每个车辆对应的解码模块进行下一轮车辆调度。

[0052] 本申请的可选实施例中,所述待派送站点与所述待接收站点一一对应;

[0053] 每个车辆对应的解码模块,还用于:

[0054] 将访问状态为已访问的站点设为无法访问的站点;

[0055] 将该车辆未访问的待接收站点所对应的待派送站点设为无法访问的站点;

[0056] 将待接收的货物量超过该车辆的剩余容量的站点设为无法访问的站点;

[0057] 删除无法访问的站点,将剩余站点中概率最大的站点,确定为该车辆去往的下一站点。

[0058] 本申请的可选实施例中,所述解码模块为基于注意力机制的神经网络。

[0059] 本申请的可选实施例中,所述的路径规划装置还包括训练单元,用于训练所述策略模型,所述训练单元包括:

[0060] 获取子单元,用于获取多组训练数据,每组所述训练数据包括每个站点的位置、所述多个站点中任意两个站点间的导航信息、每个站点待投递的货物数量、每个站点待拾取的货物数量、货物的起始地和目的地、所述多个车辆中各个车辆的容量以及车辆的数目;

[0061] 初始化子单元,初始化策略模型的参数;

[0062] 训练子单元,用于搭建策略模型,将所述多组训练数据输入所述策略模型,输出与每组训练数据对应的派送策略序列以及每个派送策略中各客户站点被选取的概率;利用所述派送策略序列计算所有车辆的总成本,以及局部解对应的局部成本;计算所述派送策略序列的评估值函数网络模型,所述评估值函数网络模型的输入为所述派送策略序列的每步局部解,输出为每步局部评估值;基于值函数更新所述策略模型和所述评估值函数;

[0063] 保存子单元,用于重复所述训练步骤,直至满足预设的训练结束条件后,保存策略模型。

[0064] 本申请实施例还提供了一种路径规划设备,包括:

[0065] 处理器;和存储器,在所述存储器中存储有计算机程序指令,

[0066] 其中,在所述计算机程序指令被所述处理器运行时,使得所述处理器执行如上所述的路径规划方法中的步骤。

[0067] 本申请实施例还提供了一种计算机可读存储介质,所述计算机可读存储介质存储

有计算机程序,所述计算机程序被处理器运行时,使得所述处理器执行如上所述的路径规划方法中的步骤。

[0068] 本申请的实施例具有以下有益效果:

[0069] 本实施例预先训练好策略模型,利用预先训练好的策略模型生成路径规划策略,能够大幅度提高路径规划策略的生成速度,由于考虑了与时间相关的站点间的行驶时间,本申请实施例能够生成更符合实际交通情况的路径规划策略,能够优化车辆的行驶路径和/或行驶时间。另外,由于采用了基于注意力机制的神经网络和采用了动态池化的全卷积网络进行编码,本申请实施例能够不受训练样本中的站点数量的限制,能够在推导时为不同数量的站点规划路径。

附图说明

[0070] 图1为本申请实施例的策略模型的一种结构示意图;

[0071] 图2为本申请实施例一车辆路径规划方法的流程示意图;

[0072] 图3为本申请实施例的策略模型的编码器的一种结构示意图;

[0073] 图4为本申请实施例的策略模型的解码器的一种结构示意图;

[0074] 图5为本申请实施例的得到策略模型的流程示意图;

[0075] 图6为本申请实施例的得到策略模型的另一流程示意图;

[0076] 图7为本申请实施例的生成路径规划策略的流程示意图;

[0077] 图8为本申请实施例的生成融合特征向量的一种网络结构示意图;

[0078] 图9和图10为本申请实施例的路径规划装置的结构框图;

[0079] 图11为本申请实施例的车辆路径规划设备的结构框图。

具体实施方式

[0080] 为使本申请的实施例要解决的技术问题、技术方案和优点更加清楚,下面将结合附图及具体实施例进行详细描述。

[0081] 本申请实施例提供一种时间依赖的路径规划方法及装置,能够扩展路径规划的适用场景,提高所生成的路径规划策略的质量。

[0082] 实施例一

[0083] 本申请实施例提供一种时间依赖的路径规划方法,用于利用预先训练的策略模型生成多个车辆访问多个站点的路径,如图1所示,本申请实施例的策略模型包括有编码器和解码器,所述多个站点包括待派送站点和待接收站点。本申请实施例中,假设各个待派送站点和待接收站点仅能被一个车辆访问一次,当待派送站点或待接收站点被某个车辆访问一次后,即表示已对该站点进行了货物派送或货物接收,此时,将该待派送站点或待接收站点标记为已访问,此后任何车辆都不再对标记为已访问的站点进行访问。另外,本申请实施例中,假设待派送站点和待接收站点是一一对应关系,待派送站点的待派送货物量与待接收站点的待接收货物量相同,即一个待接收站点的全部货物,将通过一个车辆配送到对应的待派送站点。

[0084] 如图2所示,本实施例的路径规划方法包括:

[0085] 编码步骤101,将多个站点中任意两个站点间的导航信息以及每个站点的站点属

性信息,输入至所述编码器,以生成融合特征向量;其中,所述导航信息包括一个统计时间周期内的每个时间段所对应的任意两个站点间的行驶时间,所述站点属性信息包括站点的位置、待检收的货物量、待派送的货物量以及货物的起始地和目的地,所述融合特征向量包括每个站点在每个时间段对应的特征向量。

[0086] 这里,导航信息包括多个站点中任意两个站点间的行驶时间。另外,导航信息还可以包括多个站点中任意两个站点间的行驶距离。考虑到实际路网中的交通拥堵情况是动态变化的,例如,同一个路段在不同的时间段可能有不同的行驶时间。也就是说,实际应用中,行驶时间是一种时间依赖的导航信息。为了能获得更符合实际交通情况的路径规划策略,本申请实施例可以预先获取一个统计时间周期内的每个时间段所对应的任意两个站点间的行驶时间。这里,每个时间段所对应的任意两个站点间的行驶时间,是指在该时间段中出发时该任意两个站点间的行驶时间。具体的出发时间可以是该时间段中的任意时间。

[0087] 本申请实施例中,所述统计时间周期可以是一个自然日、一个星期甚至一个月等。以自然日为例,每个时间段可以根据具体路网交通情况来划分,例如,路网中交通拥堵指数的波动范围在预设区间内的连续时间作为一个时间段,这样,同一个时间段内该路段的通行时间比较接近,同一个时间段内的该路段的通行时间可以取该时间段内的平均通行时间。当然,本申请实施例也可以将一天中的每5分钟、每半个小时、每1个小时或每个小时作为一个时间段。各个时间段的长度可以相同或不同,这样,一个统计时间周期可以划分成多个连续且互不重叠的时间段。具体划分方式本申请实施例不做具体限定。

[0088] 本申请实施例中,所述导航信息可以从外部的基于地图的导航系统获取的。行驶距离是指按照导航系统规划的两个站点之间的行驶路线的距离,而非两个站点之间的直线距离。某个时间段所对应的两个站点间的行驶时间,是指导航系统预测的车辆在该时间段出发时完成上述行驶路线所需要的时间。例如,本申请实施例可以利用上述行驶距离和/或行驶时间,生成每个时间段对应的一个 $M \times M$ 维的矩阵,所述矩阵中的元素 $C_{i,j}$ 表示该时间段对应的站点间的行驶时间和/或行驶距离,例如,元素 $C_{i,j}$ 表示站点 i 和站点 j 间的行驶时间和/或行驶距离。上述行驶距离和行驶时间对于特定的两个站点来说都是静态值。假设所述统计时间周期包括有 T 个时间段,则最终可以得到一个 $M \times M \times T$ 维的矩阵(即导航信息)。这里, M 表示所有站点(包括车辆的始发站点、待派送站点和待检收站点)的数量。

[0089] 所述站点属性信息包括站点的位置、待检收的货物量、待派送的货物量以及货物的起始地和目的地。可选的,所述站点属性信息还可以包括站点的标识,站点类型(如待派送站点或待检收站点)。所述站点的位置具体可以是站点的坐标。所述货物量具体可以是货物重量和货物体积中的至少一种。

[0090] 解码步骤102,根据每个车辆从当前站点出发时所处的当前时间段,从所述融合特征向量中提取每个站点在当前时间段对应的特征向量,并输入至所述解码器,获得所述解码器解码得到的路径规划策略,所述路径规划策略包括每个车辆访问站点的顺序。

[0091] 通过以上步骤,本申请实施例在编码生成站点的特征向量时,将站点间的时间依赖的导航信息融入到特征向量中,将该特征向量输入到解码器,从而在路径规划策略中考虑了时间依赖的导航信息的影响,所生成的路径规划策略更符合实际交通情况,从而提高了路径规划策略的质量。

[0092] 如图3所示,作为一种实现方式,本申请实施例中策略模型的编码器可以包括第一

网络、第二网络和融合模块,其中:

[0093] 所述第一网络根据所述多个站点中任意两个站点间的导航信息,生成每个站点的第一特征向量。具体的,所述第一网络将导航信息($M \times M \times T$ 维的矩阵)压缩到 $M \times K \times T$ 维的第一特征向量,其中,每一行对应每一个站点的特征向量, K 为一个固定整数。

[0094] 所述第二网络根据所述多个站点中每个站点的站点属性信息,生成每个站点的第二特征向量。具体的,所述第二网络根据每个站点的站点属性信息,学习站点间的关联信息,进而根据站点间的关联信息,生成每个站点的第二特征向量。

[0095] 所述融合模块融合每个站点的第一特征向量和第二特征向量,获得每个站点的融合特征向量。所述融合的具体方式可以包括向量拼接、向量相加和向量相乘中的一种或多种。

[0096] 可选的,所述第二网络还可以根据所述多个站点中每个站点的站点属性信息以及任意两个站点间的导航信息,生成每个站点的第二特征向量。此时,所述第二网络根据每个站点的站点属性信息以及任意两个站点间的导航信息,学习站点间的关联信息,进而根据站点间的关联信息,生成每个站点的第二特征向量。

[0097] 这里,所述第一网络可以是采用了动态池化的全卷积网络,所述第二网络可以是基于注意力机制的神经网络。通过动态池化的全卷积网络,可以适应于训练阶段和推导阶段的站点总数的变化,使得模型不再要求训练阶段和推导阶段的站点总数需要相同。

[0098] 如图4所示,作为一种实现方式,本申请实施例中策略模型的解码器可以包括交互模块、与车辆一一对应的解码模块、矛盾处理模块和策略输出模块。这里,假设一共有 N 个车辆, N 为大于或等于1的整数,在应用于多车辆路径规划时,所述 N 为大于或等于2的整数。在每轮车辆调度中:

[0099] 每个车辆对应的解码模块,根据该车辆从当前站点出发时所处的当前时间段,从所述融合特征向量中提取每个站点在当前时间段对应的特征向量,根据所提取的特征向量和该车辆所在的当前站点,计算该车辆去往各个站点的概率,确定该车辆去往的下一站点。例如,选择概率最大的站点作为该车辆去往的下一个站点。具体的,所述解码模块可以基于注意力机制的神经网络。

[0100] 所述矛盾处理模块在所有解码模块都确定了对应车辆去往的下一站点后,若存在至少两个车辆去往同一个下一站点,则选择所述至少两个车辆中的一个车辆去往该下一站点,其他车辆停留在原站点。例如,假设有3个车辆都去往站点 x ,则可以根据每个车辆去往该站点 x 的概率,选择概率最大的车辆去往该站点 x ,而剩余的2个车辆则停留在原站点,即保持在原地。当然,也可以从所述至少两个车辆中随机选择一个车辆去往该下一站点,其他车辆停留在原站点,本申请对此不做具体限定。

[0101] 所述交互模块根据各个车辆去往的下一站点,更新每个车辆访问的站点列表、各个站点的访问状态、各个车辆所在的当前站点和剩余容量、以及各个车辆从当前站点出发时所处的当前时间段。这里,根据车辆最终去往的下一站点,更新每个车辆所在的当前站点(将下一站点设置为当前站点)和站点的访问状态(将去往的站点设置为已访问);根据车辆的当前剩余容量、下一站点待派送的货物量或待接收的货物量,更新车辆的剩余容量;以及,根据去往下一个站点所需要的行驶时间,更新车辆从当前站点出发时所处的当前时间段。例如,假设预测车辆在一天内的第9个时间段(8:00-9:00)从当前站点 a 出发,去往的下

一站点为站点b,从站点a到站点b的行驶时间为1个小时,则将当前站点更新为站点b,当前时间段更新为一天内的第10个时间段(9:00-10:00)。在车辆到达站点后,若该站点为待检收站点,车辆应全部取走站点的待检收货物;若站点为待派送站点,车辆应将站点的待派送货物全部派送到站点。即站点被车辆访问后,站点的状态由未被访问更新为已访问,之后策略模型可以屏蔽该站点,不再计算该站点的备选概率值,以减少计算量。

[0102] 例如,车辆的当前剩余容量为20吨,将待检收货物量为8吨的站点分配给车辆后,车辆的剩余容量更新为12吨,待检收货物量为8吨的站点的状态由未访问变为已访问;之后策略模型将屏蔽该站点,不再计算该站点的备选概率值。比如车辆的当前剩余容量为20吨,将待派送货物量为10吨的站点分配给车辆后,车辆的剩余容量更新为30吨,待派送货物量为10吨的站点的状态由未访问变为已访问;之后策略模型将屏蔽该站点,不再计算该站点的备选概率值。

[0103] 所述策略输出模块根据更新后的各个站点的访问状态,判断所有站点是否都被访问,若是,则结束调度,输出所述路径规划策略,否则,继续由每个车辆对应的解码模块进行下一轮车辆调度。

[0104] 可选的,本申请实施例还可以将所有车辆均已返回始发站点作为结束调度的一个条件,此时,所述策略输出模块还可以根据更新后的各个站点的访问状态,判断所有站点是否都被访问,以及所有车辆始发是否都已返回始发站点,若是,则结束调度,输出所述路径规划策略,否则,继续由每个车辆对应的解码模块进行下一轮车辆调度。

[0105] 本申请实施例中,所述待派送站点与所述待检收站点一一对应。每个车辆对应的解码模块,在确定该车辆去往的下一站点时,在得到该车辆去往各个站点的概率后,可以先将访问状态为已访问的站点设为无法访问的站点;将该车辆未访问的待检收站点所对应的待派送站点设为无法访问的站点;将待检收的货物量超过该车辆的剩余容量的站点设为无法访问的站点;然后,删除无法访问的站点,将剩余站点中概率最大的站点,确定为该车辆去往的下一站点。

[0106] 举例来说,假设待派送站点A1与待检收站点A2对应,需要将待检收站点A2的货物派送到待派送站点A1;待派送站点B1与待检收站点B2对应,需要将待检收站点B2的货物派送到待派送站点B1。若所述待检收站点未被访问,则对应的待派送站点为无法访问的站点,比如在车辆未到达待检收站点B2之前,待派送站点B1为无法访问的站点;在车辆未到达待检收站点A2之前,待派送站点A1为无法访问的站点。在车辆到达待检收站点之后,对应的待派送站点为可以访问的站点。

[0107] 可以看出,本实施例中,在生成车辆去往各个站点的概率之后,在将概率最大的站点分配给当前被调度车辆之前,还需要去除无法访问的站点,无法访问的站点包括货物量超过当前被调度车辆的剩余容量的站点、对应的待检收站点未被访问的待派送站点、已经被访问过的站点等等,其中,当前所在的站点可以不包括在无法访问的站点内。

[0108] 通过以上模块,本申请实施例编码器采用采用了动态池化的全卷积网络作为第一网络,采用基于注意力机制的神经网络作为第二网络,由于采用了动态池化的全卷积网络和基于注意力机制的神经网络能够基于不同大小的输入维度,生成对应的特征向量,并在生成特征向量的过程中引入了时间依赖的导航信息以反应路网的实际通行时间,因此,本申请实施例的策略模型可以适应于参与路径规划的站点数量的变化,即训练策略模型时的

训练样本中的站点数量,不需要和实际推导时的站点数量相同,从而大大提高了策略模型的灵活性,扩展了策略模型的适用场景,使得预测得到的路径规划策略更符合路网的实际交通情况。

[0109] 在所有站点分配完毕(即都已经被访问)后,可以根据站点加入车辆路径列表的顺序,得到每一车辆的行驶路径,还可以得到每一车辆的行驶总时长,可以根据每一车辆的行驶总时长评价路径规划策略的优劣。

[0110] 一具体示例中,如表1所示,站点的信息如下:

站点标识	0	1	2	3	4	5	6	7	8	9	10
站点坐标	0.5, 0.4	3.6, 2.4	2.1, 1.3	9.2, 5.1	7.3, 0.4	8.9, 3.3	1.1, 5.0	7.8, 7.2	6.2, 9.5	2.5, 8.2	3, 5
待派送 货物量		0	0	0	0	0	3	1	4	2	1
待接收 货物量		3	1	4	2	1	0	0	0	0	0
对应的 站点标识		6	7	8	9	10	1	2	3	4	5

[0111] 表1

[0113] 其中,待派送货物量和待接收货物量的单位可以为吨,标识为0的站点也就是车辆的始发站点(仓库),标识为1-10的站点为待派送站点或待接收站点,待派送站点的待接收货物量为0,待接收站点的待派送货物量为0,站点的坐标为平面坐标系下的坐标,单位可以是千米,车辆的初始容量可以为10吨,一共有两辆车。站点6与站点1对应,需要将站点1的货物派送到站点6;站点7与站点2对应,需要将站点2的货物派送到站点7;站点8与站点3对应,需要将站点3的货物派送到站点8;站点9与站点4对应,需要将站点4的货物派送到站点9;站点10与站点5对应,需要将站点5的货物派送到站点10。该示例中,一共有11个站点(即 $M=11$),2个车辆。假设统计时间周期为一个自然日,时间段为一个自然日的24个小时。

[0114] 本实施例中,策略模型可以包括编码器和解码器,所述解码器包括有与每个车辆相对应的解码模块,通过每一车辆对应的解码模块为每一车辆计算下一步动作(去往的下一站点)的备选概率值,判断车辆是停留在原地还是去往其他站点。

[0115] 本实施例中,车辆从仓库出发,根据编码器获得的 $11 \times 24 \times d$ 维编码矩阵,及两辆车当前所处的当前时间段 $t_1=0, t_2=0$,获取两个车对应的当前时间段的每个站点的特征向量,各自为 $11 \times d$ 维,这里, d 为编码器的模型超参数。再由两个车辆对应的解码模块分别生成去往站点1-10的所有站点的访问概率。

[0116] 车辆1先进行决策,判断不会停留原地;删除无法访问的站点(站点6-10,因其货物未被收取),从其余站点中选取概率最大的站点2访问;车辆2后进行决策,判断不会停留原地;删除无法访问的站点(站点6-10,因其货物未被收取,以及站点2,因已被访问),从其余站点中选择概率最大的站点1进行访问。

[0117] 该轮调度决策中,车辆1从仓库到达站点2,捡收站点2的货物,行路路径为站点0->2;车辆2从仓库到达站点1,捡收站点1的货物,行路路径为站点0->1;更新站点1、2的访问状态为已访问;以及更新车辆1、2的当前时间段。在每轮调度决策完成后,都要根据车辆去往下一站点所需要的行驶时间,更新各个车辆的当前时间段,以及,根据车辆去往的下一站点,更新下一站点的访问状态,后文中不再赘述。因此,

[0118] 当前轮调度决策结果:车辆1(带站点2货物):站点0->2;车辆2(带站点1货物):站点0->1。

[0119] 由于仍然存在未被访问的站点,因此继续下一轮调度。车辆继续行驶,根据编码器获得的 $11 \times 24 \times d$ 维编码矩阵,及两辆车的当前时刻 $t_1=1, t_2=3$,获取两个车对应的当前时刻编码,各自为 $11 \times d$ 维,再由解码器分别生成两辆车1-10所有站点的访问概率。车辆1先进行决策,判断不会停留原地;删除无法访问的站点(站点6,8,9,10,因其货物未被该车收取,站点1,2因已被访问过),从其余站点中选取概率最大的站点7访问;车辆2后进行决策,判断不会停留原地;删除无法访问的站点(站点7,8,9,10,因其货物未被其收取,以及站点1,2,7,因已被访问过),从其余站点中选择概率最大的站点4进行访问。

[0120] 当前轮调度决策结果:车辆1(无货物):站点0->2->7;车辆2(带站点1,4货物):站点0->1->4。

[0121] 由于仍然存在未被访问的站点,因此继续下一轮调度。车辆继续行驶,根据编码器获得的 $11 \times 24 \times d$ 维编码矩阵,及两辆车的当前时刻 $t_1=4, t_2=6$,获取两个车对应的当前时刻编码,各自为 $11 \times d$ 维,再由解码器分别生成两辆车去往站点1-10的所有站点的访问概率。车辆1先进行决策,判断不会停留原地;删除无法访问的站点(站点6,8,9,10,因其货物未被该车收取,站点1,2,4,7因已被访问过),从其余站点中选取概率最大的站点3访问;车辆2后进行决策,判断会停留原地,不移动。

[0122] 当前轮调度决策结果:车辆1(带站点3货物):站点0->2->7->3,车辆2(带站点1,4货物):站点0->1->4->4。

[0123] 由于仍然存在未被访问的站点,因此继续下一轮调度。车辆继续行驶,根据编码器获得的 $11 \times 24 \times d$ 维编码矩阵,及两辆车的当前时刻 $t_1=7, t_2=6$,获取两个车对应的当前时刻编码,各自为 $11 \times d$ 维,再由解码器分别生成两辆车去往站点1-10的所有站点的访问概率。车辆1先进行决策,判断不会停留原地;删除无法访问的站点(站点6,9,10,因其货物未被该车收取,站点1,2,3,4,7因已被访问过),从其余站点中选取概率最大的站点8访问;车辆2后进行决策,判断不会停留原地;删除无法访问的站点(站点9,10,因其货物未被其收取,以及站点1,2,3,4,7,8,因已被访问过),从其余站点中选择概率最大的站点6进行访问。

[0124] 当前轮调度决策结果:车辆1(无货物):站点0->2->7->3->8;车辆2(带站点4货物):站点0->1->4->4->6。

[0125] 由于仍然存在未被访问的站点,因此继续下一轮调度。车辆继续行驶,根据编码器获得的 $11 \times 24 \times d$ 维编码矩阵,及两辆车的当前时刻 $t_1=10, t_2=11$,获取两个车对应的当前时刻编码,各自为 $11 \times d$ 维,再由解码器分别生成两辆车去往站点1-10的所有站点的访问概率。车辆1先进行决策,判断不会停留原地;删除无法访问的站点(站点10,因其货物未被该车收取,站点1,2,3,4,6,7,8因已被访问过),从其余站点中选取概率最大的站点5访问;车辆2后进行决策,判断不会停留原地;删除无法访问的站点(站点10,因其货物未被其收取,

以及站点1,2,3,4,6,7,8,,因已被访问过),从其余站点中选择概率最大的9进行访问。

[0126] 当前轮调度决策结果:车辆1(带站点5货物):站点0->2->7->3->8->5;车辆2(无货物):站点0->1->4->4->6->9。

[0127] 由于仍然存在未被访问的站点,因此继续下一轮调度。车辆继续行驶,根据编码器获得的 $11 \times 24 \times d$ 维编码矩阵,及两辆车的当前时刻 $t_1=12$, $t_2=14$,获取两个车对应的当前时刻编码,各自为 $11 \times d$ 维,再由解码器分别生成两辆车去往站点1-10的所有站点的访问概率。车辆1先进行决策,判断不会停留原地;删除无法访问的站点(站点1,2,3,4,5,6,7,8,9因已被访问过),从其余站点中选取概率最大的站点8访问;车辆2后进行决策,判断不会停留原地;删除无法访问的站点(站点1,2,3,4,5,6,7,8,9,10因已被访问过),无站点可前往,选择返回原点。

[0128] 当前轮调度决策结果:车辆1(无货物):站点0->2->7->3->8->5->10;车辆2(带站点4货物):站点0->1->4->4->6->9->0。

[0129] 由于仍然存在未被访问的站点,因此继续下一轮调度。车辆继续行驶,车辆继续行驶,根据编码器获得的 $11 \times 24 \times d$ 维编码矩阵,及两辆车的当前时刻 $t_1=15$, $t_2=14$,获取两个车对应的当前时刻编码,各自为 $11 \times d$ 维,再由解码器分别生成两辆车去往站点1-10的所有站点的访问概率。车辆1先进行决策,判断不会停留原地;删除无法访问的站点(站点1,2,3,4,5,6,7,8,9,10因已被访问过),只能返回原点;车辆2后进行决策,判断不会停留原地;删除无法访问的站点(站点1,2,3,4,5,6,7,8,9,10因已被访问过),无站点可前往,选择返回原点。

[0130] 当前轮调度决策结果:车辆1(无货物):站点0->2->7->3->8->5->10->0;车辆2(带站点4货物):站点0->1->4->4->6->9->0->0。

[0131] 由于所有站点都已访问,且所有车辆均已位于起始站点,因此结束调度,最终输出每个车辆的路径规划策略如下:

[0132] 车辆1:站点0->2->7->3->8->5->10->0;

[0133] 车辆2:站点0->1->4->4->6->9->0->0。

[0134] 本实施例中,预先训练好策略模型,利用预先训练好的策略模型生成路径规划策略,能够大幅度提高路径规划策略的生成速度,由于考虑了基于时间变化的导航信息(行驶时间),本申请实施例能够生成更符合实际路况的路径规划策略,尽量使车辆的行驶路径最短、费用最小、时间最少。另外,由于采用了基于注意力机制的神经网络和采用了动态池化的全卷积网络进行编码,本申请实施例能够不受训练样本中的站点数量的限制,能够在推导时为不同数量的站点规划路径。另外,由于是利用预先训练的策略模型进行路径规划,相比于传统的启发式算法搜索最优解的实现方式,本申请实施例能够大大减少路径规划时间。

[0135] 本申请实施例中,所述策略模型为利用训练数据训练得到。路径规划方法还包括训练所述策略模型的步骤,如图5所示,训练所述策略模型的步骤包括:

[0136] 获取步骤201,获取多组训练数据,每组所述训练数据包括每个站点的位置、所述多个站点中任意两个站点间的导航信息、每个站点待投递的货物数量、每个站点待拾取的货物数量、货物的起始地和目的地、所述多个车辆中各个车辆的容量以及车辆的数目;所述导航信息包括一个统计时间周期内的每个时间段所对应的任意两个站点间的行驶时间。

[0137] 初始化步骤202,初始化策略模型的参数;

[0138] 训练步骤203,搭建策略模型,将所述多组训练数据输入所述策略模型,输出与每组训练数据对应的派送策略序列(即各个车辆依次访问的各个客户站点的序列)以及每个派送策略序列中各客户站点被选取的概率;利用所述派送策略序列计算所有车辆的总成本(这里,成本可以采用路程和/或行驶时间来表征),以及局部解对应的局部成本(所述局部解为每轮调度结束所得到的解);可选地,还可以计算所述派送策略序列的评估值函数网络模型。可选地,为Critic网络,所述评估值函数网络模型的输入为所述派送策略序列的每步局部解,输出为每步局部评估值;基于值函数更新所述策略模型和所述评估值函数;

[0139] 其中,所述基于值函数更新所述策略模型和所述评估值函数包括:

[0140] 计算所述策略模型和所述评估值函数的损失函数L;

[0141] 对所述损失函数L求梯度;

[0142] 利用梯度下降法更新所述策略模型和所述评估值函数的参数。

[0143] 一具体示例中,可以根据以下公式计算所述策略模型和所述评估值函数的损失函数L;

$$[0144] \quad L = L_{actor} + L_{critic}$$

$$[0145] \quad L_{actor} = - \sum_{t=1}^T \log \pi_{\theta}(a_t | s_t) [G(s_t, a_t) - V_{\theta}^{\pi}(s_t)]$$

$$[0146] \quad L_{critic} = L_{\delta}(G, V_{\theta}^{\pi})$$

[0147] 其中, $\pi_{\theta}(a_t | s_t)$ 表示每步对应的概率; $G(s_t, a_t)$ 表示分步的局部成本(如局部路程); $V_{\theta}^{\pi}(s_t)$ 表示每步得到的局部值; s_t 表示每次解码时的网络状态,具体包括各个车辆的位置、车辆已装载的货物量、各个客户站点的访问状态等; a_t 表示每次解码时所选取的客户站点; θ 表示策略模型及评估值函数网络的待训练参数; L_{critic} 表示对应的Huber损失函数(HuberLoss); δ 为HuberLoss的子参数; T 表示解码的总步数。

[0148] 步骤204:重复所述训练步骤,直至满足预设的训练结束条件后,进入步骤205以保存策略模型,例如,判断训练步骤的重复次数达到预设次数,如果否,转向步骤203;如果是,转向步骤205;所述预设次数可以根据需要设定。

[0149] 步骤205:保存策略模型。

[0150] 在经过上述步骤201-205得到策略模型后,保存策略模型,之后即可利用保存的策略模型生成路径规划策略,能够大大提升路径规划策略的生成速度,并使得路径规划策略更符合实际交通情况,从而减少配送成本,提高配送效率。

[0151] 图6给出了本申请实施例训练策略模型的一种总体流程图,图7则进一步给出了图6中的生成路径规划策略的步骤的实现流程。其中,如图6所示,本申请实施例在模型训练过程中,计算损失函数时,还可以加入车辆未能在期望时间内将货物派送至目标站点的超时惩罚项。

[0152] 本申请实施例的编码器,可以采用时域增强的transformer结构。图8给出了本申请实施例可以基于每个站点的第一特征向量和站点间的导航信息生成所述融合特征向量的一种网络结构示意图,其中:

[0153] 该网络的输入向量包括 x 和 ω , 其中, x 为每个站点的第一特征向量,具有 $M \times K \times T$ 的

维度; ω 为所述多个站点中任意两个站点间的导航信息矩阵(第二特征矩阵), 具有 $M \times M \times T$ 的维度; 该网络的输出向量 y 为所述每个站点的融合特征向量, 具有 $M \times K \times T$ 的维度。在该网络的网络结构中, 输入的第一特征向量经三个线性层(Linear), 分别获得 V, Q, K 三个分量特征。其中, Q, K 分量特征相乘和缩放处理后再进行指数归一化处理, 然后, 与导航信息矩阵融合处理(可相加, 相乘或相联合等), 融合处理所得的特征再与 V 分量特征相乘, 从而得到自注意力特征。此若干层可合称为矩阵增强的自注意力层。此后, 自注意力特征经线性层和前馈层的处理, 获得最终的融合特征向量输出。

[0154] 实施例二

[0155] 本申请实施例还提供了一种时间依赖的路径规划装置, 用于利用预先训练的策略模型生成多个车辆访问多个站点的路径, 所述策略模型包括有编码器和解码器, 所述多个站点包括待派送站点和待接收站点, 如图9所示, 本实施例的路径规划装置包括:

[0156] 信息输入单元601, 用于信息输入单元, 用于将多个站点中任意两个站点间的导航信息以及每个站点的站点属性信息, 输入至所述编码器, 以生成融合特征向量; 其中, 所述导航信息包括一个统计时间周期内的每个时间段所对应的任意两个站点间的行驶时间, 所述站点属性信息包括站点的位置、待接收的货物量、待派送的货物量以及货物的起始地和目的地, 所述融合特征向量包括每个站点在每个时间段对应的特征向量;

[0157] 策略输出单元602, 用于根据每个车辆从当前站点出发时所处的当前时间段, 从所述融合特征向量中提取每个站点在当前时间段对应的特征向量, 并输入至所述解码器, 获得所述解码器解码得到的路径规划策略, 所述路径规划策略包括每个车辆访问站点的顺序。

[0158] 如实施例一所述的, 本申请实施例的策略模型的编码器包括第一网络、第二网络和融合模块, 其中:

[0159] 所述第一网络, 用于根据所述多个站点中任意两个站点间的导航信息, 生成每个站点的第一特征向量;

[0160] 所述第二网络, 用于根据所述多个站点中每个站点的站点属性信息, 生成每个站点的第二特征向量;

[0161] 所述融合模块, 用于融合每个站点的第一特征向量和第二特征向量, 获得所述融合特征向量。

[0162] 可选的, 所述第二网络, 还用于根据所述多个站点中每个站点的站点属性信息以及任意两个站点间的导航信息, 生成每个站点的第二特征向量。

[0163] 可选的, 所述第一网络为采用了动态池化的全卷积网络, 所述第二网络为基于注意力机制的神经网络。

[0164] 本申请实施例的策略模型的解码器包括交互模块、与车辆一一对应的解码模块、矛盾处理模块和策略输出模块;

[0165] 每个车辆对应的解码模块, 用于在每轮车辆调度中, 根据该车辆从当前站点出发时所处的当前时间段, 从所述融合特征向量中提取每个站点在当前时间段对应的特征向量, 根据所提取的特征向量和该车辆所在的当前站点, 计算该车辆去往各个站点的概率, 确定该车辆去往的下一站点;

[0166] 所述矛盾处理模块, 用于在每轮车辆调度中, 在所有解码模块都确定了对应车辆

去往的下一站点后,若存在至少两个车辆去往同一个下一站点,则选择所述至少两个车辆中的一个车辆去往该下一站点,其他车辆停留在原站点;

[0167] 所述交互模块,用于在每轮车辆调度中,根据各个车辆去往的下一站点,更新每个车辆访问的站点列表、各个站点的访问状态、各个车辆所在的当前站点和剩余容量、以及各个车辆从当前站点出发时所处的当前时间段;

[0168] 所述策略输出模块,用于根据更新后的各个站点的访问状态,判断所有站点是否都已经被访问,若是,则将全部车辆调回始发站点,并输出所述路径规划策略,否则,继续由每个车辆对应的解码模块进行下一轮车辆调度。

[0169] 这里,所述待派送站点与所述待接收站点一一对应;

[0170] 每个车辆对应的解码模块,还用于在每轮调度中:

[0171] 将访问状态为已访问的站点设为无法访问的站点;

[0172] 将该车辆未访问的待接收站点所对应的待派送站点设为无法访问的站点;

[0173] 将待接收的货物量超过该车辆的剩余容量的站点设为无法访问的站点;

[0174] 删除无法访问的站点,将剩余站点中概率最大的站点,确定为该车辆去往的下一站点。

[0175] 可选的,所述解码模块为基于注意力机制的神经网络。

[0176] 如图10所示,本申请实施例提供的路径规划装置,还包括:

[0177] 训练单元603,用于训练所述策略模型。

[0178] 可选的,所述训练单元包括以下子单元:

[0179] 获取子单元,用于获取多组训练数据,每组所述训练数据包括每个站点的位置、所述多个站点中任意两个站点间的导航信息、每个站点待投递的货物数量、每个站点待拾取的货物数量、货物的起始地和目的地、所述多个车辆中各个车辆的容量以及车辆的数目;

[0180] 初始化子单元,初始化策略模型的参数;

[0181] 训练子单元,用于搭建策略模型,将所述多组训练数据输入所述策略模型,输出与每组训练数据对应的派送策略序列以及每个派送策略中各客户站点被选取的概率;利用所述派送策略序列计算所有车辆的总成本,以及局部解对应的局部成本;计算所述派送策略序列的评估值函数网络模型,所述评估值函数网络模型的输入为所述派送策略序列的每步局部解,输出为每步局部评估值;基于值函数更新所述策略模型和所述评估值函数;

[0182] 保存子单元,用于重复所述训练步骤,直至满足预设的训练结束条件后,保存策略模型。

[0183] 可选的,所述训练子单元,还用于计算所述策略模型和所述评估值函数的损失函数 L ;对所述损失函数 L 求梯度;利用梯度下降法更新所述策略模型和所述评估值函数的参数。

[0184] 实施例三

[0185] 本申请实施例还提供了一种时间依赖的路径规划设备50,用于利用预先训练的策略模型生成多个车辆访问多个站点的路径,所述策略模型包括有编码器和解码器,所述多个站点包括待派送站点和待接收站点。如图11所示,包括:

[0186] 处理器52;和

[0187] 存储器54,在所述存储器54中存储有计算机程序指令,

[0188] 其中,在所述计算机程序指令被所述处理器运行时,使得所述处理器52执行以下步骤:

[0189] 编码步骤,将多个站点中任意两个站点间的导航信息以及每个站点的站点属性信息,输入至所述编码器,以生成融合特征向量;其中,所述导航信息包括一个统计时间周期内的每个时间段所对应的任意两个站点间的行驶时间,所述站点属性信息包括站点的位置、待检收的货物量、待派送的货物量以及货物的起始地和目的地,所述融合特征向量包括每个站点在每个时间段对应的特征向量;

[0190] 解码步骤,根据每个车辆从当前站点出发时所处的当前时间段,从所述融合特征向量中提取每个站点在当前时间段对应的特征向量,并输入至所述解码器,获得所述解码器解码得到的路径规划策略,所述路径规划策略包括每个车辆访问站点的顺序。

[0191] 本实施例中,预先训练好策略模型,利用预先训练好的策略模型生成站点的备选概率值,根据备选概率值的大小为车辆分配站点,生成路径规划策略,能够大幅度提高路径规划策略的生成速度。

[0192] 进一步地,如图11所示,多车辆路径规划设备50还包括网络接口51、输入设备53、硬盘55、和显示设备56。

[0193] 上述各个接口和设备之间可以通过总线架构互连。总线架构可以是包括任意数量的互联的总线和桥。具体由处理器52代表的一个或者多个中央处理器(CPU),以及由存储器54代表的一个或者多个存储器的各种电路连接在一起。总线架构还可以将诸如外围设备、稳压器和功率管理电路等之类的各种其它电路连接在一起。可以理解,总线架构用于实现这些组件之间的连接通信。总线架构除包括数据总线之外,还包括电源总线、控制总线和状态信号总线,这些都是本领域所公知的,因此本文不再对其进行详细描述。

[0194] 所述网络接口51,可以连接至网络(如因特网、局域网等),从网络中获取相关数据,并可以保存在硬盘55中。

[0195] 所述输入设备53,可以接收操作人员输入的各种指令,并发送给处理器52以供执行。所述输入设备53可以包括键盘或者点击设备(例如,鼠标,轨迹球(trackball)、触感板或者触摸屏等。

[0196] 所述显示设备56,可以将处理器52执行指令获得的结果进行显示。

[0197] 所述存储器54,用于存储操作系统运行所必须的程序和数据,以及处理器52计算过程中的中间结果等数据。

[0198] 可以理解,本申请实施例中的存储器54可以是易失性存储器或非易失性存储器,或可包括易失性和非易失性存储器两者。其中,非易失性存储器可以是只读存储器(ROM)、可编程只读存储器(PROM)、可擦除可编程只读存储器(EPROM)、电可擦除可编程只读存储器(EEPROM)或闪存。易失性存储器可以是随机存取存储器(RAM),其用作外部高速缓存。本文描述的装置和方法的存储器54旨在包括但不限于这些和任意其它适合类型的存储器。

[0199] 在一些实施方式中,存储器54存储了如下的元素,可执行模块或者数据结构,或者他们的子集,或者他们的扩展集:操作系统541和应用程序542。

[0200] 其中,操作系统541,包含各种系统程序,例如框架层、核心库层、驱动层等,用于实现各种基础业务以及处理基于硬件的任务。应用程序542,包含各种应用程序,例如浏览器(Browser)等,用于实现各种应用业务。实现本申请实施例方法的程序可以包含在应用程序

542中。

[0201] 关于策略模型的具体结构,请参考实施例一和实施例二的描述,此处不再赘述。

[0202] 本申请上述实施例揭示的方法可以应用于处理器52中,或者由处理器52实现。处理器52可能是一种集成电路芯片,具有信号的处理能力。在实现过程中,上述方法的各步骤可以通过处理器52中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器52可以是通用处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现成可编程门阵列(FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件,可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器54,处理器52读取存储器54中的信息,结合其硬件完成上述方法的步骤。

[0203] 可以理解的是,本文描述的这些实施例可以用硬件、软件、固件、中间件、微码或其组合来实现。对于硬件实现,处理单元可以实现在一个或多个专用集成电路(ASIC)、数字信号处理器(DSP)、数字信号处理设备(DSPD)、可编程逻辑设备(PLD)、现场可编程门阵列(FPGA)、通用处理器、控制器、微控制器、微处理器、用于执行本申请所述功能的其它电子单元或其组合中。

[0204] 对于软件实现,可通过执行本文所述功能的模块(例如过程、函数等)来实现本文所述的技术。软件代码可存储在存储器中并通过处理器执行。存储器可以在处理器中或在处理器外部实现。

[0205] 进一步地,处理器52还用于执行训练所述策略模型的步骤,其中,训练所述策略模型的步骤包括:

[0206] 获取步骤,获取多组训练数据,每组所述训练数据包括每个站点的位置、所述多个站点中任意两个站点间的导航信息、每个站点待投递的货物数量、每个站点待拾取的货物数量、货物的起始地和目的地、所述多个车辆中各个车辆的容量以及车辆的数目;

[0207] 初始化步骤,初始化策略模型的参数;

[0208] 训练步骤,搭建策略模型,将所述多组训练数据输入所述策略模型,输出与每组训练数据对应的派送策略序列以及每个派送策略中各客户站点被选取的概率;利用所述派送策略序列计算所有车辆的总成本,以及局部解对应的局部成本;计算所述派送策略序列的评估值函数网络模型,所述评估值函数网络模型的输入为所述派送策略序列的每步局部解,输出为每步局部评估值;基于值函数更新所述策略模型和所述评估值函数;

[0209] 重复所述训练步骤,直至满足预设的训练结束条件后,保存策略模型。

[0210] 进一步地,处理器52具体用于计算所述策略模型和所述评估值函数的损失函数L;

[0211] 对所述损失函数L求梯度;

[0212] 利用梯度下降法更新所述策略模型和所述评估值函数的参数。

[0213] 进一步地,处理器52具体用于根据分配给当前被调度车辆的站点的待派送货物量或待检收货物量以及当前被调度车辆的当前容量更新当前被调度车辆的剩余容量。

[0214] 进一步地,所述待派送站点与所述待检收站点一一对应,若所述待检收站点未被

访问,则对应的待派送站点为无法访问的站点,处理器52具体用于删除无法访问的站点,将前往剩余站点和停留原地中备选概率值最大的动作分配给当前被调度车辆。

[0215] 实施例四

[0216] 本申请实施例还提供了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器运行时,使得所述处理器执行以下步骤:

[0217] 编码步骤,将多个站点中任意两个站点间的导航信息以及每个站点的站点属性信息,输入至所述编码器,以生成融合特征向量;其中,所述导航信息包括一个统计时间周期内的每个时间段所对应的任意两个站点间的行驶时间,所述站点属性信息包括站点的位置、待检收的货物量、待派送的货物量以及货物的起始地和目的地,所述融合特征向量包括每个站点在每个时间段对应的特征向量;

[0218] 解码步骤,根据每个车辆从当前站点出发时所处的当前时间段,从所述融合特征向量中提取每个站点在当前时间段对应的特征向量,并输入至所述解码器,获得所述解码器解码得到的路径规划策略,所述路径规划策略包括每个车辆访问站点的顺序。

[0219] 本实施例中,预先训练好策略模型,利用预先训练好的策略模型生成站点的备选概率值,根据备选概率值的大小为车辆分配站点,生成路径规划策略,能够大幅度提高路径规划策略的生成速度。

[0220] 进一步地,所述计算机程序被处理器运行时,还使得所述处理器执行以下步骤:训练所述策略模型的步骤,训练所述策略模型的步骤包括:

[0221] 获取步骤,获取多组训练数据,每组所述训练数据包括每个站点的位置、所述多个站点中任意两个站点间的导航信息、每个站点待投递的货物数量、每个站点待拾取的货物数量、货物的起始地和目的地、所述多个车辆中各个车辆的容量以及车辆的数目;

[0222] 初始化步骤,初始化策略模型的参数;

[0223] 训练步骤,搭建策略模型,将所述多组训练数据输入所述策略模型,输出与每组训练数据对应的派送策略序列以及每个派送策略中各客户站点被选取的概率;利用所述派送策略序列计算所有车辆的总成本,以及局部解对应的局部成本;计算所述派送策略序列的评估值函数网络模型,所述评估值函数网络模型的输入为所述派送策略序列的每步局部解,输出为每步局部评估值;基于值函数更新所述策略模型和所述评估值函数;

[0224] 重复所述训练步骤,直至满足预设的训练结束条件后,保存策略模型。

[0225] 进一步地,所述计算机程序被处理器运行时,还使得所述处理器执行以下步骤:计算所述策略模型和所述评估值函数的损失函数L;

[0226] 对所述损失函数L求梯度;

[0227] 利用梯度下降法更新所述策略模型和所述评估值函数的参数。

[0228] 进一步地,所述计算机程序被处理器运行时,还使得所述处理器执行以下步骤:根据分配给当前被调度车辆的站点的待派送货物量或待检收货物量以及当前被调度车辆的当前容量更新当前被调度车辆的剩余容量。

[0229] 进一步地,所述待派送站点与所述待检收站点一一对应,若所述待检收站点未被访问,则对应的待派送站点为无法访问的站点,所述计算机程序被处理器运行时,还使得所述处理器执行以下步骤:删除无法访问的站点,将前往剩余站点和停留原地中备选概率值最大的动作分配给当前被调度车辆。

[0230] 以上是本申请的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本申请所述原理的前提下,还可以作出若干改进和润饰,这些改进和润饰也应视为本申请的保护范围。

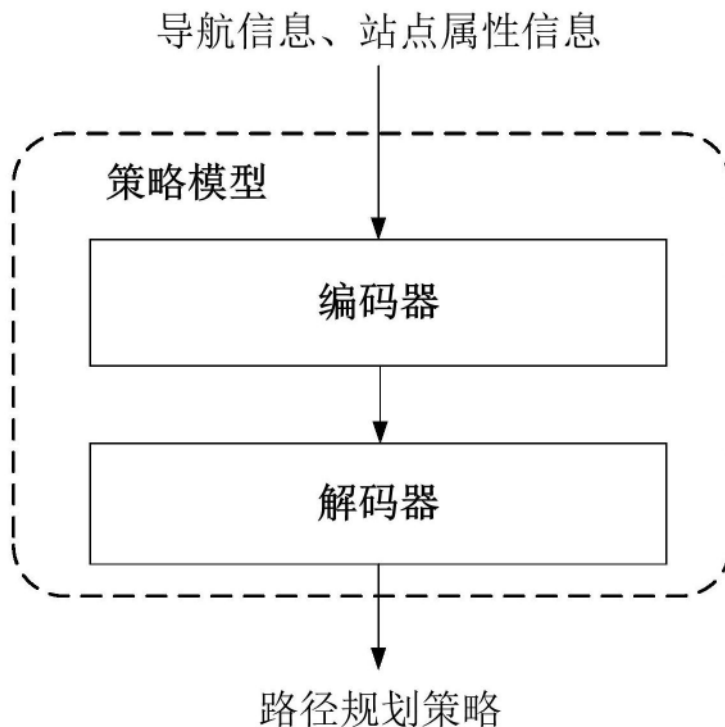


图1

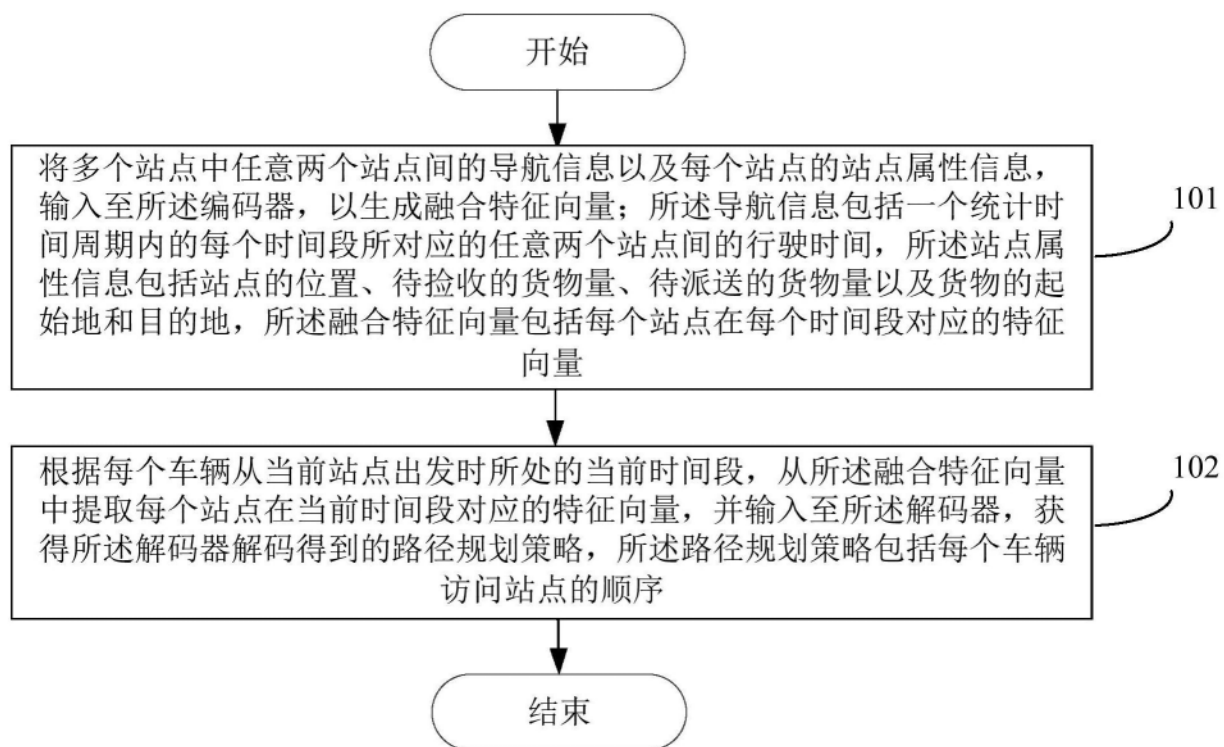


图2

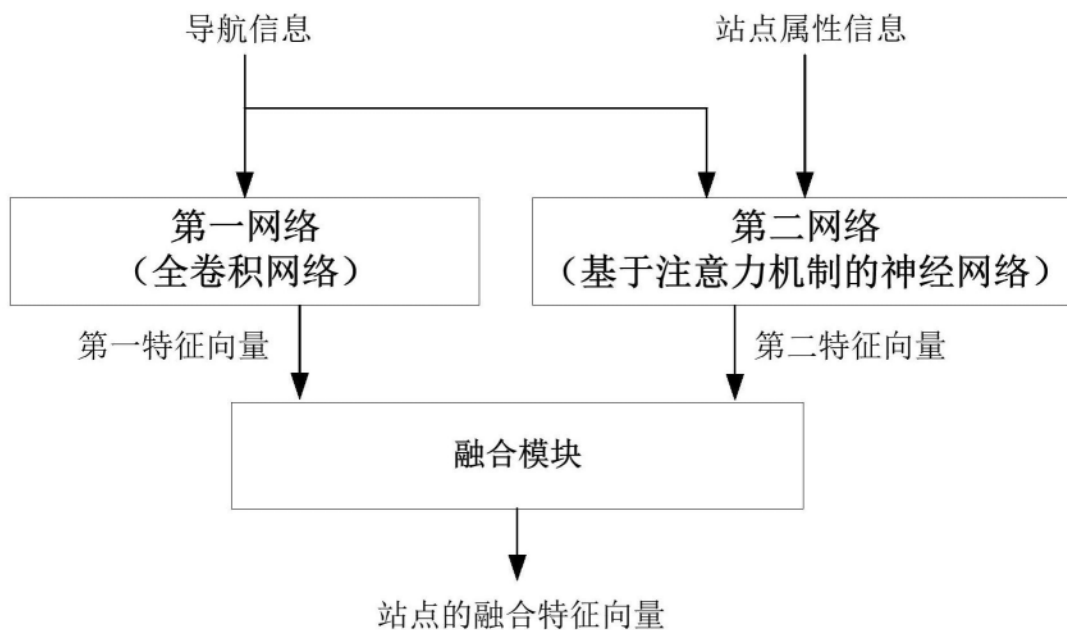


图3

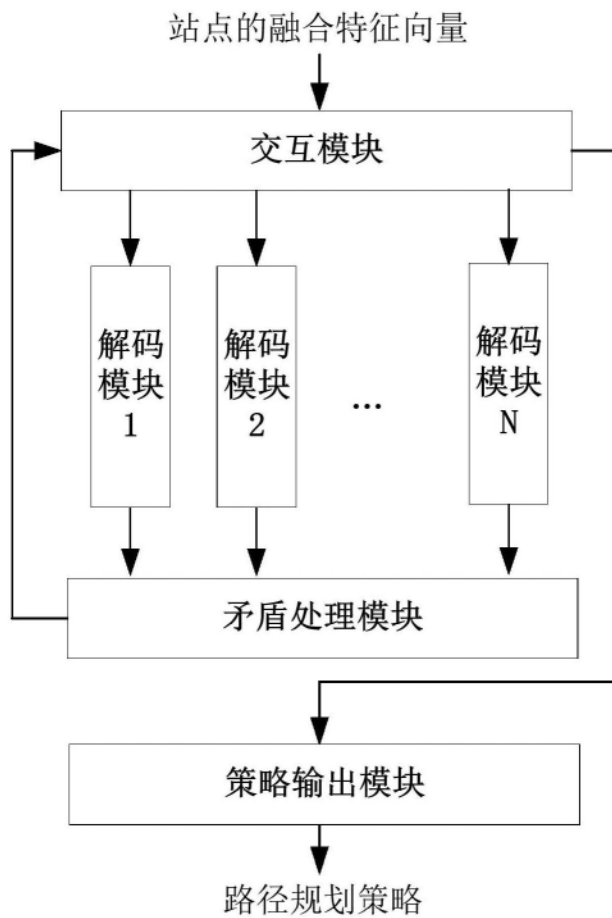


图4

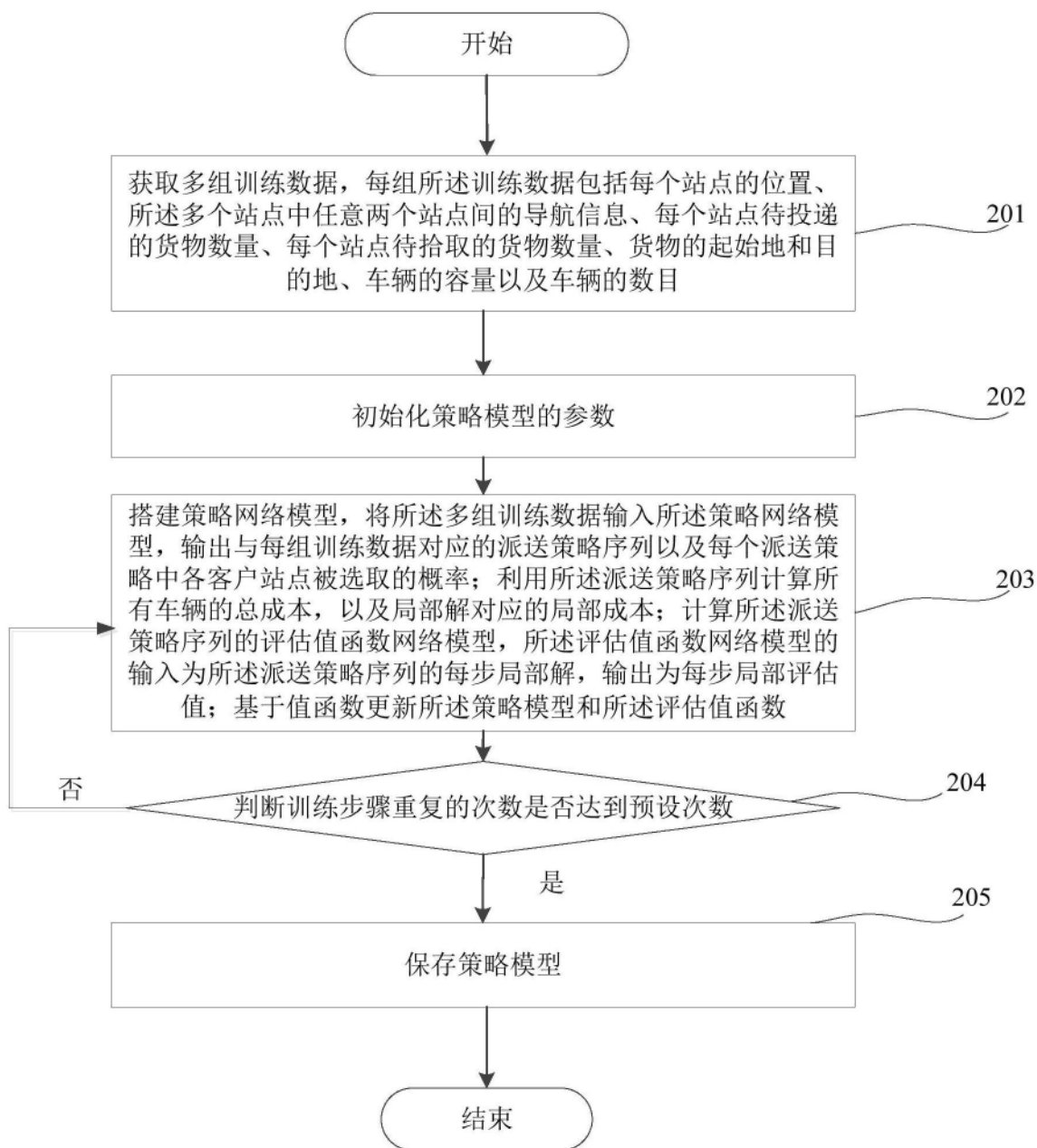


图5

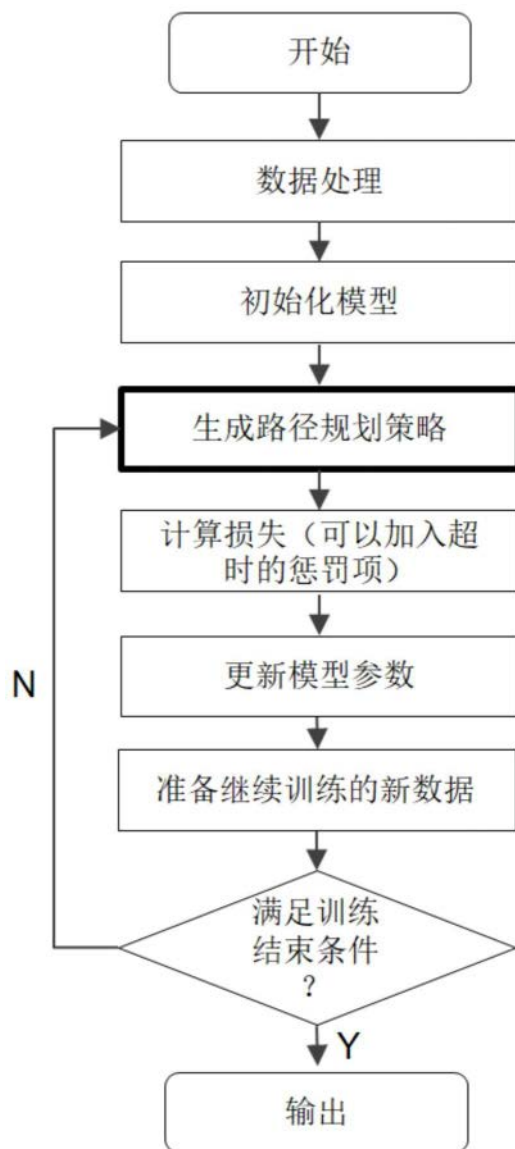


图6

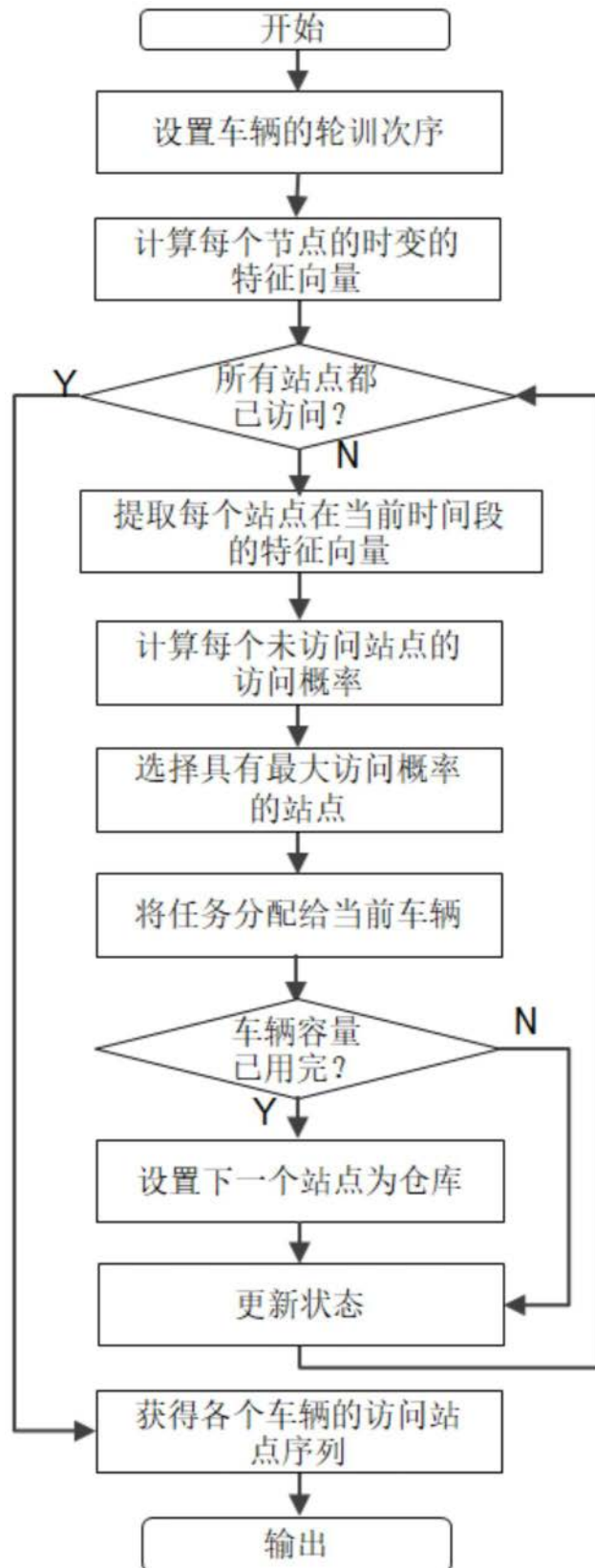


图7

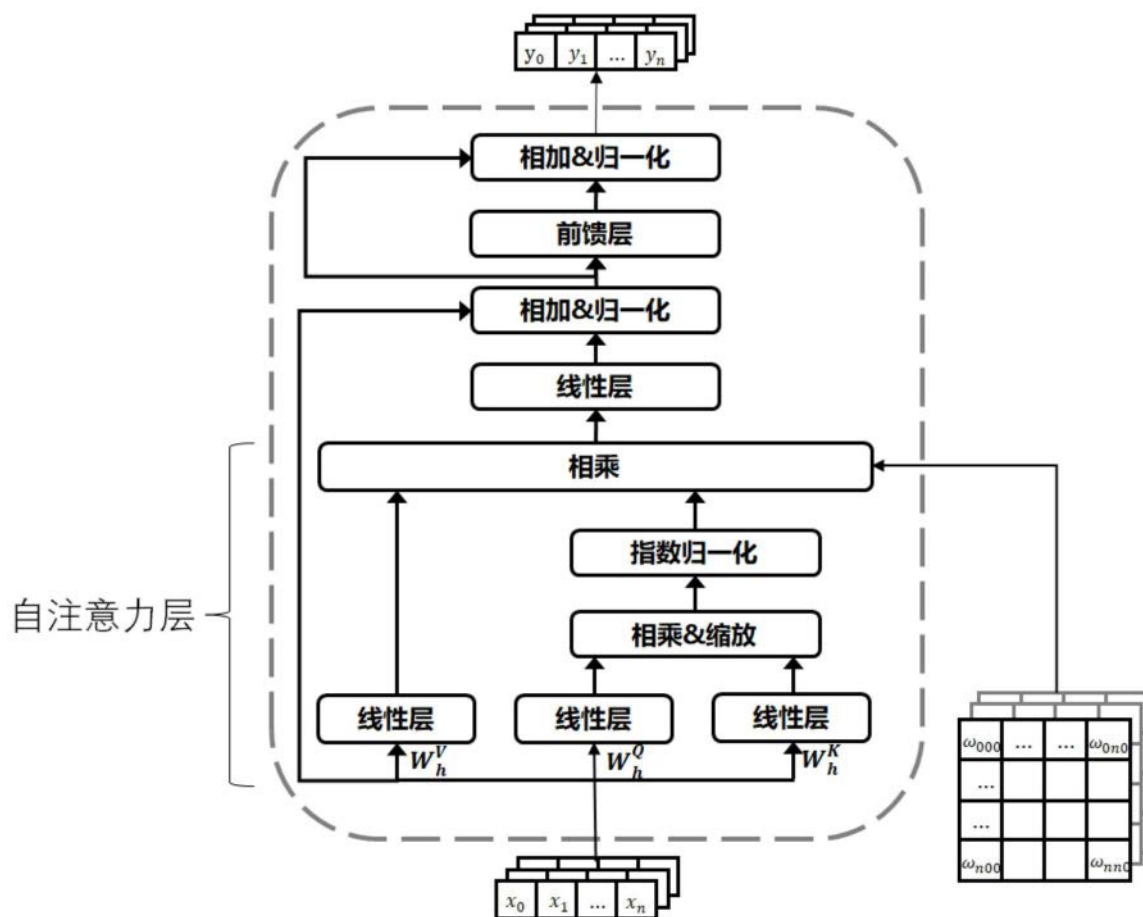


图8

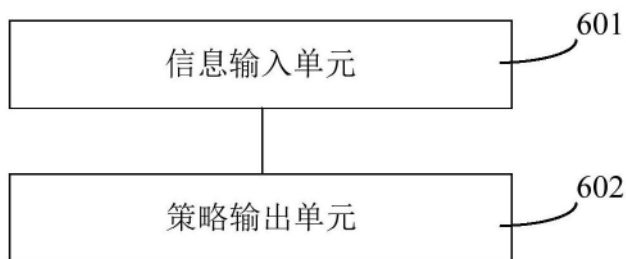


图9

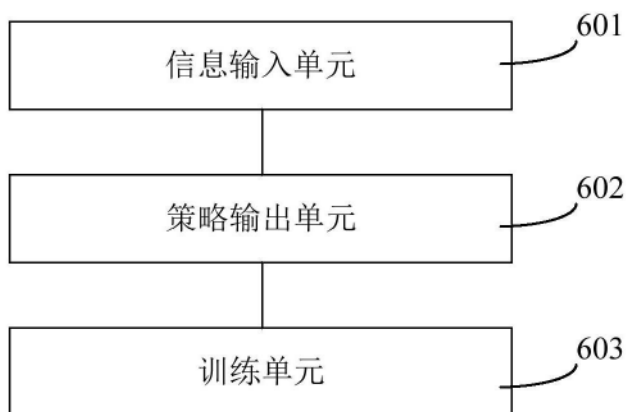


图10

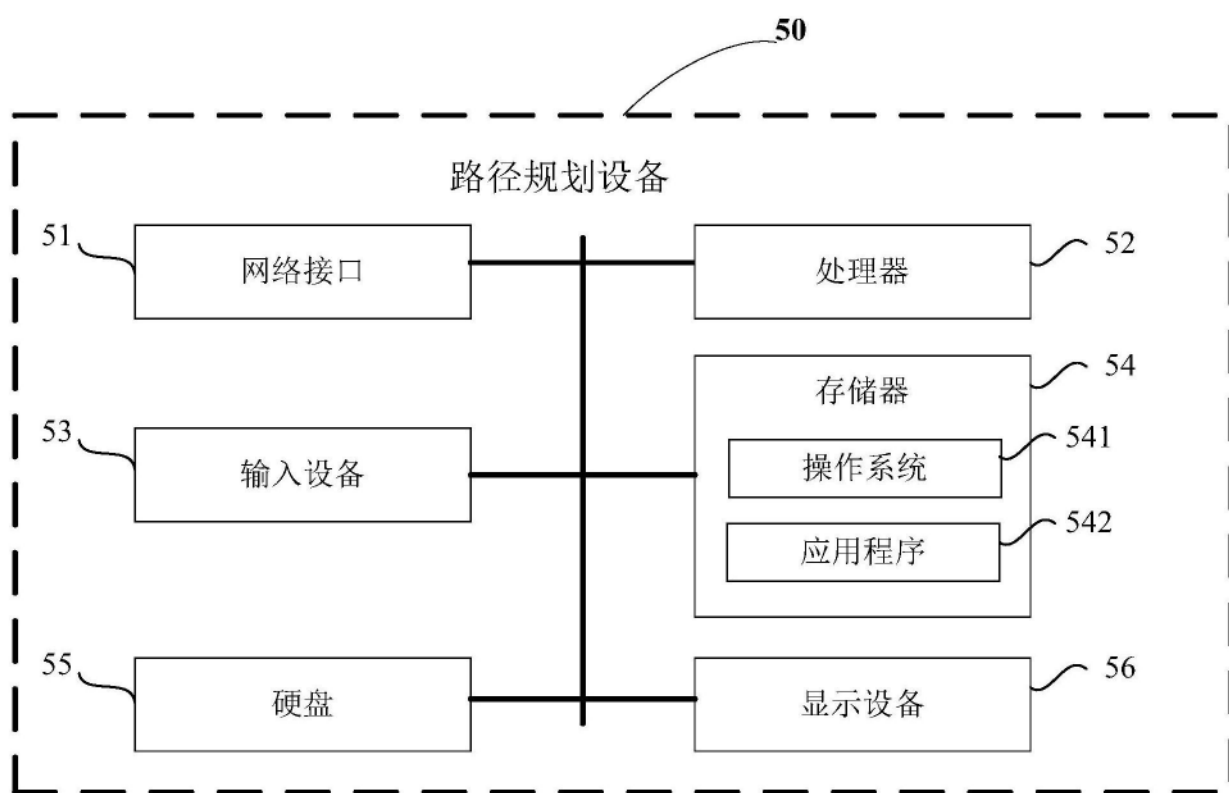


图11