

引文格式:叶浩宇,涂伟,叶贺辉,等.基于深度强化学习的电动出租车运营优化[J].测绘学报,2020,49(12):1630-1639. DOI:10.11947/j. AGCS.2020.20190516.  
YE Haoyu, TU Wei, YE Hehui, et al. Deep reinforcement learning based electric taxi service optimization[J]. Acta Geodaetica et Cartographica Sinica, 2020, 49(12): 1630-1639. DOI: 10.11947/j. AGCS.2020.20190516.

## 基于深度强化学习的电动出租车运营优化

叶浩宇<sup>1</sup>, 涂伟<sup>2,3,4,5</sup>, 叶贺辉<sup>6</sup>, 麦可<sup>2,3,4</sup>, 赵天鸿<sup>2,4</sup>, 李清泉<sup>1,2,3,4</sup>

1. 武汉大学测绘遥感信息工程国家重点实验室,湖北武汉 430079; 2. 深圳大学建筑与城市规划学院城市空间信息工程系,广东深圳 518060; 3. 人工智能与数字经济广东省实验室(深圳),广东深圳 518060; 4. 广东省城市空间信息工程重点实验室,广东深圳 518060; 5. 自然资源部大湾区地理环境监测重点实验室,广东深圳 518060; 6. 闽江学院软件学院,福建福州 350108

## Deep reinforcement learning based electric taxi service optimization

YE Haoyu<sup>1</sup>, TU Wei<sup>2,3,4,5</sup>, YE Hehui<sup>6</sup>, MAI Ke<sup>2,3,4</sup>, ZHAO Tianhong<sup>2,4</sup>, LI Qingquan<sup>1,2,3,4</sup>

1. State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China; 2. Department of Urban Informatics, School of Architecture and Urban Planning, Shenzhen University, Shenzhen 518060, China; 3. Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen University, Shenzhen 518060, China; 4. Guangdong Key Laboratory of Urban Informatics, Shenzhen University, Shenzhen 518060, China; 5. Key Laboratory for Geo-Environmental Monitoring of Great Bay Area, MNR, Shenzhen 518060, China; 6. Software College, Minjiang University, Fuzhou 350108, China

**Abstract:** Electric taxis have been demonstrated with the promotion of electric vehicles. Compared with internal combustion engine vehicles, electric taxis spend more time in recharging, which reduces the taxi drivers' intention to use. Reinforcement learning is applicable to the sequential decision-making process of taxis driver. This paper presents the double deep Q-learning network (DDQN) model to simulate the operation of electric taxis. According to the real-time state of taxis, DDQN will choose the optimal actions to execute. After training, we obtain a global optimal electric taxi service strategy, and finally optimize the taxi service. Using real-world taxi travel data, an experiment is conducted in Manhattan Island in New York City, USA. Results show that, comparing with the baseline methods, DDQN reduces the waiting time for charging and the rejection rate by 70% and 53%, respectively. Taxi drives' income are finally increased by about 7%. Moreover, the results of model parameter sensitivity analysis indicate that the charge speed and the number of vehicles have greater impact on drives' income than the battery capacity. When the charging rate reaches 120 kW, electric taxis achieve the best performance. The government should build more fast charging station to improve the revenue of electric taxis.

**Key words:** deep reinforcement learning; electric taxi; DDQN; taxi service strategies

**Foundation support:** The National Key Research and Development Program of China (No. 2019YFB2103104); The Natural Science Foundation of Guangdong Province (No. 2019A1515011049); The Basic Research Projects of Shenzhen Technology Innovation Commission (No. JCYJ20170412105839839)

**摘要:**作为公共交通的重要组成部分,电动出租车对电动车推广具有重要的示范意义。相较于燃油出租车,电动出租车需要耗费更多充电时间,降低了出租车司机使用意愿,全面推广面临较大阻力。强化学习方法方兴未艾,适用于出租车运营的顺序决策过程。基于强化学习,本文构建双深度 Q 学习网络(double deep Q-learning network, DDQN)模型模拟电动出租车的运行。根据出租车的实时状态选择并执行最优载客、充电、空驶和等待等动作,通过训练得到全局最优的电动出租车运营策略,实现电动出租车运营智能优化。利用美国纽约市曼哈顿岛的出租车出行数据进行试验。结果表明:相较于简单的电动出租车运营模式,DDQN 优化策略最高将充电等待时长降低 70%,拒载率降低 53%,司机的载客收入提高 7%。相对于电池容量,充电速率和车辆总数对出租车运营收入影响更大,当充电速率达到

120 kW 时,电动出租车达到最佳的运营表现,政府在推广电动出租车的过程中应当建设更多高速率的充电站以提升电动出租车的运营表现。

**关键词:**深度强化学习;电动出租车;DDQN;出租车运营

**中图分类号:**P208

**文献标识码:**A

**文章编号:**1001-1595(2020)12-1630-10

**基金项目:**国家重点研发计划(2019YFB2103104);广东省自然科学基金(2019A1515011049);深圳市科技创新委员会基础研究(JCYJ20170412105839839)

交通运输是能源消耗和碳排放的主要来源之一<sup>[1]</sup>,推行节能减排已成为当下的国际共识<sup>[2]</sup>。作为城市公共交通的重要组成部分,低排放、高效的电动出租车日渐受到关注。随着移动互联网和人工智能的蓬勃发展<sup>[3]</sup>,网约车革新了传统出租车运营方式<sup>[4]</sup>。智能网联的电动出租车有望提供更加高效、节能的出行服务<sup>[5-8]</sup>,缓解交通拥堵、减少碳排放<sup>[9-10]</sup>。许多城市和地区已经制定或实施了出租车电动化方案:英国伦敦将于 2020 年实现 9000 辆插电混合动力出租车上路运营的目标<sup>[11]</sup>;深圳市于 2019 年成为世界首个公交车、出租车基本完成电动化的城市。

与传统燃油出租车相比,电动出租车在充电过程中需耗费 0.5~12 h,有效运营时间减少,影响司机收入,不利于大规模推广。为减少充电过程对出租车司机有效运营时间对收入的影响,国内外学者已开展相关研究,主要分为 3 类。第 1 类方法是充电设施空间优化布局。文献[12]建模出行时空需求、电动出租车与充电站交互过程,以最大化充电站和出租车的服务能力、减少电动出租车的充电等待时间为目标,提出了充电站选址优化模型。然而,充电设施改善需要较长时间。司机的充电决策不仅取决于充电站的分布,也受到司机主观意识影响。第 2 类研究关注充电设施负载均衡调度策略。文献[13]采用分散控制策略均衡高峰-峡谷用电时间上的充电需求,降低排队等待时间。文献[14]建立了一个基于预约的调度系统以分配充电请求,减少电动汽车司机在充电站的等待时间。第 3 类研究为出租车司机充电方案推荐优化。文献[15]通过挖掘电动出租车的历史充电事件,建立实时充电站推荐系统,使得充电等待时间降低 50%。文献[16]利用大规模出租车 GPS 轨迹数据提出电动出租车实时路径推荐系统,使得司机收入高于 76.2%的汽油出租车司机。

上述方法着眼于充电需求-充电设施供应,没有考虑司机其他行为(如:载客、充电)对司机收益的影响。例如出租车司机充电时错过产生订单的

高峰期,导致后续较低的运营收益。同时,上述研究主要针对电动出租车当时的状态进行决策,没有考虑决策之后产生的影响:当出租车载客后,下车点可能是一个乘客稀少的区域,进而导致了潜在的空载返程成本。在长时间运营过程上,司机的连续决策构成一个长时间的序列,只考虑当前状态的决策通常不是全局最优。为了优化整个决策序列,需综合考虑每个决策对全局的影响。

强化学习(reinforcement learning)<sup>[17]</sup>是基于马尔可夫决策过程(Markov decision process, MDP)建立决策模型,在与环境的交互过程中学习经验,优化决策的一种机器学习方法。强化学习适用于连续决策过程,在博弈游戏<sup>[18]</sup>和机器人自学习<sup>[19]</sup>等领域表现优异。针对燃油出租车,文献[20]利用强化学习估计出租车司机在不同位置的最优选择,利用北京的历史轨迹数据进行试验,表明该方法能有效提高司机利润。其他基于强化学习的出租车司机策略推荐研究<sup>[21-22]</sup>也证实了强化学习在出租车行为决策优化的良好性能。但上述研究多针对传统燃油出租车,尚未考虑电动出租车的复杂的载客、充电、空驶和等待过程。

本文面向电动出租车复杂运营过程,构建适应电动出租车的复杂运营特性的 MDP 框架,提出优化电动出租车运营决策的双深度 Q 学习网络(double deep Q-learning network, DDQN),为电动出租车推荐长期价值最大化的策略,优化电动出租车运营。研究框架如图 1 所示,包含数据预处理、DDQN 模型训练和在线测试 3 个部分。本框架可用于传统、网约或自动驾驶电动出租车的运营决策优化,促进电动出租车的推广,实现更加高效、节能和环保的出行。

## 1 基于深度强化学习算法的电动出租车运营优化

### 1.1 电动出租车运营优化

本文基于马尔可夫决策过程,构建深度强化学习模型,估计不同决策行为(包括载客、充电、空

驶和原地等待)对电动出租车运营的长期影响,寻找司机收益最大化的决策行为,如图2所示。马尔可夫决策过程可对系统连续决策中的不确定性进行建模,广泛应用于序列化决策优化问题。MDP由状态 $s$ 、动作 $a$ 、状态转移 $T$ 和奖励 $r$ ,4个基本元素组成。本文采用MDP描述出租车的运营过程。首先,将电动出租车的状态定义为 $s = [t \ l_i \ b_i \ o_i]$ ,其中, $t$ 表示时间, $l_i$ 和 $b_i$ 分别为所处位置与电量, $o_i$ 为载客所获得的收入。可供电动出租车选择的动作集合为 $a \in \{p, c, e, w\}$ ,其中 $p$ 表示前往乘客出发点进行载客; $c$ 代表选择附近某个充电站进行充电; $e$ 和 $w$ 分别为选择空驶巡航和原地等待乘客。

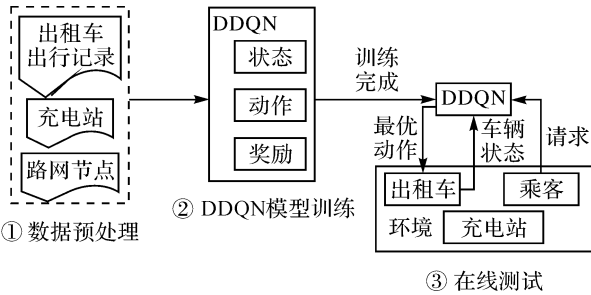


图1 基于深度强化学习的电动出租车运营优化

Fig.1 Deep reinforcement learning based electric taxi service optimization

位置 $l_i$ 前往载客点进行载客,行驶至落客点 $l_{i+1}$ ;②充电,出租车前往充电站充电,充电前可能需要排队等待;③空驶,前往附近的任一地点 $l_{i+1}$ ;④原地等待,出租车执行动作的电量消耗 $E$ 包括行驶耗电和附加耗电,与其行驶速度 $v$ 、距离 $d$ 与时长 $T$ 相关

$$E = \beta(\alpha_1 v^2 + \alpha_2 v + \alpha_3) \times d + \frac{\mu T}{60} \quad (1)$$

式中, $\beta$ 表征驾驶习惯对电量消耗的影响,附加耗电系数 $\mu$ 与当地温度相关,相关参数取值与文献[23]一致。

电动出租车运营决策过程即是根据其所处状态 $s = [t \ l_i \ b_i \ o_i]$ ,选择最优的动作 $a_t$ 。相应地,状态发生转移 $T = \{s_t \rightarrow s_{t+1}\}$ ,演变为下一个状态 $s_{t+1} = [t+1 \ l_{t+1} \ b_{t+1} \ o_{t+1}]$ 。

根据电动出租车运营过程中的收入和损失,本文设置奖励函数为出租车执行相应动作产生的收益。如表1所示,每种动作具有相应的奖励函数 $r$ 。

表1 不同动作的奖励函数

Table.1 The rewards of actions

动作 $a$	奖励函数 $r$
载客 $p$	$r_{od} - T_o C$
充电 $c$	$-P_{elec} - T_c C$
空驶 $e$	$-T_e C$
等待 $w$	$-T_w C$

表1中, $C$ 是 $t$ 时刻出租车在 $l$ 地点的单位时间期望载客收益,本文用其代表出租车未载客时的损失,可根据历史出行数据计算得到。 $t$ 时刻,位于 $l$ 地点的出租车选择载客动作 $p$ ,奖励值为满足相应出行需求的收入 $r_{od}$ 与前往载客点时长 $T_o$ 内的损失之和;选择进行充电 $c$ 的奖励值等于充电收费 $P_{elec}$ 与该动作耗时 $T_c$ 内的预期损失之和;空驶 $e$ 和原地等待 $w$ 所获得的奖励则分别与执行当次动作所产生的空驶时长 $T_e$ 和等待时长 $T_w$ 负相关。

本研究采用累积衰减函数定义电动出租车决策带来的奖励。对于一辆电动出租车一段时间内的连续决策和状态转移过程,其一系列的状态、动作和奖励可表示为状态序列 $[s_t \ s_{t+1} \ s_{t+2} \ \dots \ s_T]$ 、动作序列 $[a_t \ a_{t+1} \ a_{t+2} \ \dots \ a_T]$ 和相应的奖励序列 $[r_t \ r_{t+1} \ r_{t+2} \ \dots \ r_T]$ 后,其在状态 $s_t$ 下执行动作 $a_t$ 可获得的未来回报 $R_t$ 定义为



图2 电动出租车运营决策

Fig.2 Decision for electric taxi drivers

时刻 $t$ 下,给定任意一辆位于位置 $l_i$ 并具有电量 $b_i$ 的电动出租车,可从4个可能的动作中进行选择,如图2所示:①载客,如有分配到订单,从



$$\mathbb{R}_t = \sum_{k=0}^{T-t} \lambda^k r_{t+k} \quad (2)$$

式中,折扣因子  $\lambda$  代表模型对于后续发生的奖励值的重视程度,会对未来的奖励进行不同程度的衰减。

正式地,本文问题可以定义为:给定任意一辆电动出租车的状态  $s$ ,寻找具有最大未来回报  $\mathbb{R}$  的动作  $a$ 。本文引入深度强化学习,对电动出租车不同决策带来的奖励进行学习,顾及每个决策对全局的影响,进而得到最大化  $\mathbb{R}$  的动作。

## 1.2 深度强化学习

强化学习是机器学习的重要分支,其特点是无需提前给定解决问题的策略,根据环境采取行动,通过在线自学习获得策略的改进<sup>[24]</sup>。强化学习包括智能体、环境、动作、状态、奖励等要素,如图 3 所示,智能体根据自身的状态  $s_t$ ,选择执行相应的动作  $a_t$ ,并改变自身状态。与此同时,智能体所处环境会对动作的优劣作出正面或负面的评价,即奖励  $r_t$ 。强化学习的核心是试错(trial-and-error),智能体在执行动作后环境所给予的正向或负向奖励的刺激下,逐步形成对奖励的预期,产生能获得最大收益的最优行为策略<sup>[17]</sup>。在图 3 的循环过程中,智能体不断进行在线自学习,最终其选择的动作应当使得奖励值不断增长。

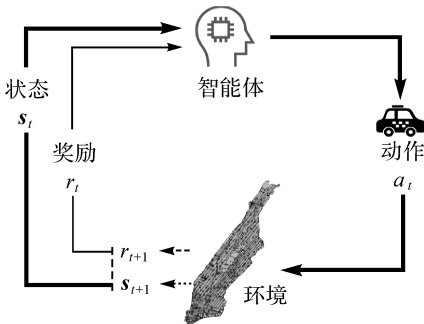


图 3 强化学习模型

Fig.3 The reinforcement learning model

强化学习算法的自学习过程通过计算状态、动作和奖励三要素,并迭代更新目标函数完成,直至目标函数收敛完成自学习过程,以达到最优控制。为实现未来收益最大化,电动出租车运营的目标函数,即最优动作-价值  $Q$  函数如下

$$Q(s, a) = \max_{\pi} \mathbb{E}[\mathbb{R}_t | s_t = s, a_t = a, \pi] \quad (3)$$

式(3)表示  $Q$  函数在状态  $s$  下选择回报值  $R_t$  最大的动作  $a$ ,其中  $\pi$  代表出租车运营的策略,  $\mathbb{E}$

表示奖励总和  $\mathbb{R}_t$  的期望。式(3)中的未来奖励值总和  $R_t$  未知,一般通过 Bellman 公式<sup>[25]</sup>进行迭代求解。通过 Bellman 公式,式(3)可简化为从动作集中选择某一个动作  $a_{t+1}$  最大化下一时刻  $Q(s_{t+1}, a_{t+1})$ ,进而最大化未来的回报  $R_t = r_t + \lambda Q(s_{t+1}, a_{t+1})$ ,即未知的最优动作-价值等价于当前时刻奖励值  $r_t$  与下一时刻的  $Q$  函数之和

$$Q(s, a) = \mathbb{E}_{S' \sim \epsilon} \left[ \begin{array}{l} \mathbb{R}_t = r_t + \\ \lambda \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) \end{array} \middle| s, a \right] \quad (4)$$

基于式(4),可采用  $Q$ -learning 算法<sup>[26]</sup>,使用时序差分<sup>[27]</sup>(temporal-difference learning)迭代计算不同状态下执行相应动作的  $Q$  函数,进行目标函数更新。利用下一时刻的  $Q$  函数值  $Q(s_{t+1}, a_{t+1})$  计算时序差分目标  $Y = r_t + \lambda \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})$ ,通过缩小当前  $Q$  函数值  $Q(s_t, a_t)$  与  $Y$  之间的差距,使其渐进收敛到目标值。其中,  $Q$  函数的迭代更新公式如下

$$Q(s, a) = Q(s_t, a_t) + \alpha (r_t + \lambda \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)) \quad (5)$$

式中,  $\alpha$  为学习率,控制算法学习迭代的速度。式(5)通过逼近时序差分目标  $Y = r_t + \lambda \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})$  的方法对  $Q$  值进行迭代更新。与监督学习不同,强化学习无须提供有标签数据进行训练,而是将  $Y$  作为“标签”,通过不断逼近  $Y$  使得  $Q$  函数最终收敛。此特性即为强化学习的自学习特征。

传统  $Q$ -learning 算法使用查找表存储  $Q$  函数值,不适用于连续状态和行为空间,在较大离散空间下需要极大的计算资源。针对电动出租车运营的复杂特性,本文使用深度神经网络  $Q(s, a; \theta)$  对  $Q$  函数进行近似,称为深度  $Q$  网络(deep  $Q$ -learning network, DQN),其中  $\theta$  表示神经网络参数<sup>[18]</sup>。深度  $Q$  网络的参数随着训练过程不断更新,使得神经网络逼近的目标,即深度  $Q$  网络计算得到的时序差分目标  $Y$  不断变动,不利于  $Y$  收敛。为使  $Y$  保持相对稳定,本文建立双深度  $Q$  网络<sup>[28]</sup>(DDQN)模型进行训练

$$Y^{DDQN} \equiv r_t + \lambda Q(s_{t+1}, \arg \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}; \theta^1); \theta^2) \quad (6)$$

DDQN 模型包含两个神经网络:估值神经网络  $DQN^1$  和目标神经网络  $DQN^2$ 。估值神经网络  $DQN^1$  用于选择下个时刻的最佳动作  $a_{t+1} = \arg \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}; \theta^1)$ ,使得  $s_{t+1}$  所对应的  $Q$  值最大,其参数  $\theta^1$  不断更新。目标神经网络  $DQN^2$  用于评估所选取的下一时刻动作  $a_{t+1}$  的价

值  $Q(s_{t+1}, a_{t+1}; \theta^2)$ , 得到时序差分目标  $Y = r_t + \lambda Q(s_{t+1}, a_{t+1}; \theta^2)$ , 其参数  $\theta^2$  在间隔时间  $\tau$  后更新参数  $\theta^2 = \theta^1$ , 使得  $Y$  在时间  $\tau$  内保持稳定。通过双深度  $Q$  网络, 实现动作选择与动作价值评估的分离, 保持一段时间内神经网络收敛目标  $Y$  的稳定, 模型训练过程更加快速且易收敛。

基于 DDQN 模型, 定义神经网络损失函数  $L(\theta)$  如下, 并采用小批量梯度下降法在每次迭代中更新神经网络参数<sup>[18]</sup>

$$L_i(\theta_i) = \mathbb{E}_{s, a \sim \rho(\cdot, \cdot)} [(Y_i^{\text{DDQN}} - Q(s, a; \theta_i))^2] \quad (7)$$

### 1.3 基于 DDQN 模型优化电动出租车运营

基于上述 DDQN 模型学习电动出租车运营的动作-价值网络, 进行电动出租车运营决策。如图 4 所示, 初始化神经网络与经验存储队列后, 通过不断递增时间步  $t$ , DDQN 模型将不同涌现的出行需求分配至距离乘客最近的车辆。出租车利用训练后的动作-价值网络判断是否可以接受该次出行请求, 或者随机性进行决策。如果 10 min 内未有车辆接单, 则该次出行未被满足。

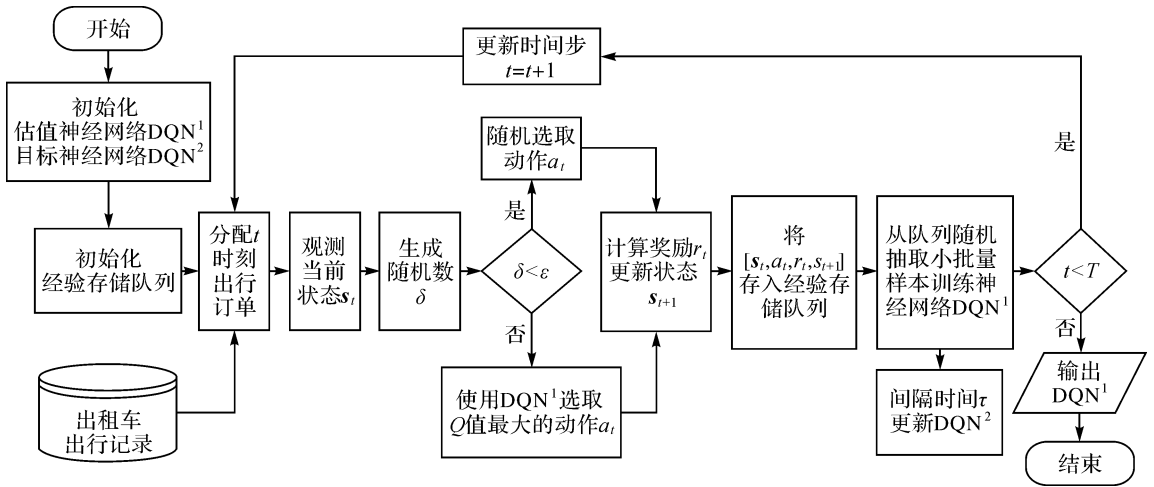


图 4 基于双深度  $Q$  学习网络的电动出租车运营优化自学习算法

Fig.4 Double deep  $Q$ -learning network based self-learning for the electric taxi operation

为避免陷入局部最优, 本文通过  $\epsilon$ -贪婪策略选择动作对环境进行探索, 即以  $1-\epsilon$  的概率选择  $Q$  值最大的动作, 以  $\epsilon$  的概率随机选取动作, 并得到相应的奖励  $r_t$ , 车辆状态转换到下一个时刻的  $s_{t+1}$ 。  $\epsilon$  值越大, 随机性越强。为加速探索过程, 本文设置较大的  $\epsilon$  初值, 并随着训练次数增加而减少  $\epsilon$ , 确保在训练后期动作选择的确定性。公式如下

$$\epsilon = P_1 + (P_0 - P_1) \exp(-kn) \quad (8)$$

式中,  $P_0$  和  $P_1$  分别为初始和结束的探索概率; 参数  $k$  控制  $\epsilon$  递减的速率;  $n$  为训练次数。为减少训练样本间的关联性, 避免模型只关注近期产生的样本, 本文采用经验回放 (experience replay) 机制, 将上述样本  $[s_t, a_t, r_t, s_{t+1}]$  保存到定长的经验存储队列中<sup>[29]</sup>, 从中随机选择小批量的样本作为神经网络的输入进行训练, 从而使得训练样本近似达到独立分布, 提高神经网络训练的

稳定性。

## 2 试验与分析

### 2.1 研究区域与试验数据

研究区域为美国纽约市曼哈顿岛, 如图 5 所示。曼哈顿岛人口密度高, 出租车出行需求巨大。全面推广电动出租车能够减少该地区化石能源消耗和温室气体排放<sup>[30]</sup>。

本文使用纽约市计程车委员会 (Taxi and Limousine Commission) 所收集的出租车出行数据 (origin destination, OD), 模拟真实的出行需求, 将其用于模型训练与测试。该数据包括载、落客时间与地点、出行距离和收费等字段, 无出租车 ID 字段, 如表 2 所示。出租车出行数据覆盖 2016 年 5 月 2 日至 13 日中的 10 个工作日, 利用 5 月 2 日—5 月 6 日的数据进行 DDQN 训练, 5 月 9 日—5 月 13 日的数据进行测试。充电站分

布采用美国能源部公布的最新的充电站数据,由于曼哈顿岛上城区暂无充电站,为了模拟提供全岛的电动出租车服务,在该区域适当人工增设若干充电站,最终共包含了 258 个充电站,如图 5 所示。

表 2 出租车出行数据示例

Tab.2 Example of taxi order

载客时间	落客时间	出行距离 /km	收费 / \$	上车点 ID	下车点 ID
07:00	07:06	2.69	10.55	5724	4262
08:10	08:13	1.45	6.8	803	1614
17:15	17:20	2.25	9.73	2149	3726



图 5 曼哈顿的充电站分布

Fig.5 The distribution of charging stations in Manhattan

## 2.2 试验参数设置

本文试验中的电动出租车的车辆参数如下: 电池容量 60 kWh,续航里程 364 km。参考曼哈顿地区的出租车保有量,电动出租车的初始数量设置为 6000 辆。充电站内充电桩的充电功率设定为 40 kW,充电量随着充电时长线性增加。为了模拟出租车司机充电决策的随机性,设定出租车充电电量超过一定数值后随机结束充电,本文设定该数值为 80%。电动出租车在  $k$ -近邻的充电站中选择充电等待时长最短的充电站进行充电。

DDQN 模型超参数包括折扣因子  $\lambda$ 、学习速率  $\alpha$ 、双神经网络更新时间  $\tau$  和小批量训练样本数  $n$ 。通过反复试验,进行参数调优,最后选择参数  $\lambda=0.9, \alpha=0.002, \tau=6000, n=64$ 。

为了验证算法性能,本文设计了距离优先和时间优先两种策略进行比较。距离优先策略下电

动出租车前往距离最近的充电站进行充电。时间优先策略下电动出租车前往等待时间最短的充电站进行充电。

## 2.3 试验结果与分析

### 2.3.1 DDQN 结果分析

采用上述参数,进行 DDQN 模型训练,模型训练损失变化如图 6 所示。随着训练时长的增加,模型损失波动较大,但整体呈现下降的趋势,最终收敛于 1000 左右,模型训练完成。

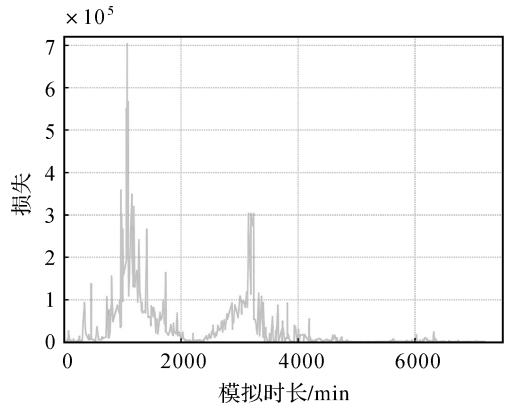


图 6 损失函数变化

Fig.6 Change of the cost function

进行 DDQN 模型测试,分析电动出租车运营优化结果,如表 3 第 1 行所示。试验结果表明,基于 DDQN 的电动出租车运营辅助下,电动出租车平均每天载客 9.76 h,行驶 138.31 km,获得收入 595.10 \$。每辆电动出租车平均每天充电 1.44 h,需等待 1.00 h;花费 9.43 h 进行空驶寻客,并在出行需求较多的地区原地等待乘客,日均时长达到 2.37 h。

图 7 进一步给出了基于 DDQN 推荐优化策略的电动出租车运营优化结果分布。在 DDQN 的优化下,95%的电动出租车日均收入高于 200 \$,平均值 595.10 \$,部分电动出租车日均收入超过 1000 \$。载客里程和空驶里程也具有类似分布。绝大部分电动出租车平均每天需进行 1—2 次的充电。整体上看,充电时长有两个高峰,分别为 1.3 h 和 1.9 h。DDQN 模型能够很好地平衡载客、充电、等待和空驶过程。图 7(e) 显示超过 50%的出租车充电等待时长为 0,使得平均充电等待时长大幅下降。DDQN 模型还能减少无意义的空驶,在部分出行需求较少区域让电动出租车进行等待,部分车辆的日均等待时间超过 2 h,

如图 7(f)所示。

图 8 给出出租车出行记录、出行拒载率和充电等待时间分布。可以发现:曼哈顿出租车出行需求集中在中城和下城。基于 DDQN 推荐优化策略,约有 3.66%出行需求未被电动出租车满足,绝大部分的区域内的拒载率低于 5%,在曼哈顿

岛中部 3 个区域内拒载率较高。图 8(c)表明在上城区、中城区北部,下城区南部出行充电等待时间较长,DDQN 优化策略倾向于在出行需求较多的地区减少充电时长、增加载客里程,在需求较少的区域则选择充电,避免空载。

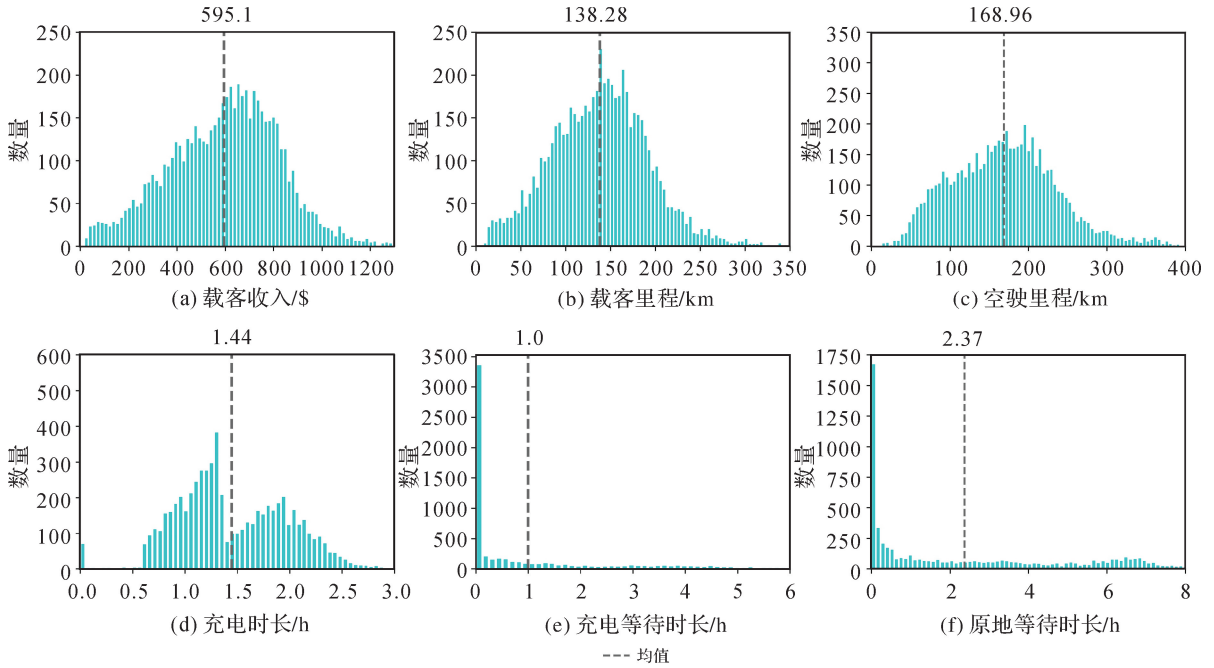


图 7 基于 DDQN 的电动出租车日平均运营结果统计

Fig.7 Statistic of the DDQN driven electric taxi's operation results

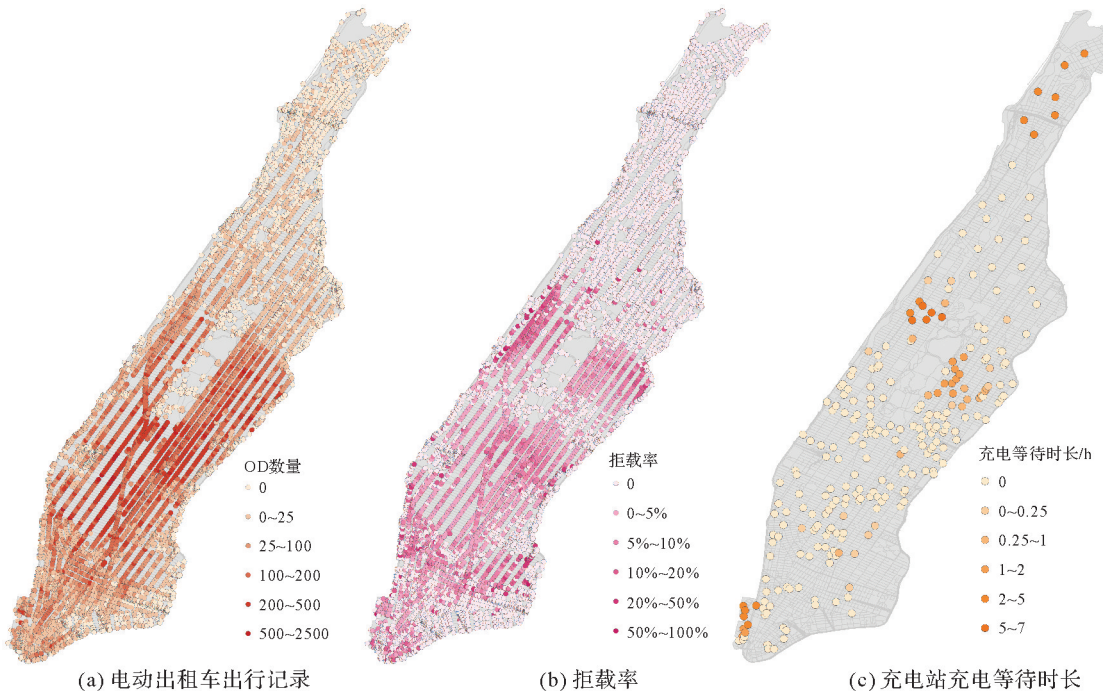


图 8 电动出租车运营结果的空间分布

Fig.8 Spatial distribution of the electric taxi's operation results



表 3 基于 DDQN、距离优先和时间优先策略的电动出租车日平均运营优化结果

Tab.3 The results of the DDQN, the distance-first and the time-first strategy

策略	载客			充电		空驶		原地等待
	收入/\$	里程/km	时长/h	时长/h	等待时长/h	里程/km	时长/h	时长/h
DDQN	595.10	138.31	9.76	1.44	1.00	168.96	9.43	2.37
时间优先	574.97	131.97	9.41	1.69	1.26	180.79	11.64	0
距离优先	557.99	128.41	9.09	1.39	3.08	177.93	10.44	0

2.3.2 结果对比

表 3 比较了 DDQN 模型结果与距离优先和时间优先策略。DDQN 的优化策略效果更好。距离优先和时间优先策略的总拒载率分别为 7.71% 和 5.87%，DDQN 模型的总拒载率为 3.66%，较两种简单策略分别降低 53% 和 38%。DDQN 模型下的电动出租车日均载客收入更高，达到 595.10 \$，分别较距离优先、时间优先策略高 6.65% 和 3.50%。表 3 中 DDQN 的平均载客里程和时长较高，与收入相符，表明 DDQN 的优化策略更加高效。DDQN 综合考虑出租车所处时空位置、电量灵活进行载客-空驶-等待-充电决策，能够有效实现充电站之间的负载均衡，避免电动出租车在某些时刻集中选择某一充电站，充电等

待时长(1 h)分别为距离优先策略(3.08 h)和时间优先策略(1.26 h)的 30% 和 79%。时间优先策略在充电时选择附近等待时间最短的充电站，使其充电等待时长较距离优先策略大幅下降，与 DDQN 接近。但时间优先策略未能将节省的时间充分用于载客，而是进行充电和空驶，因而与 DDQN 的载客收入仍存在较大差距。经过训练后，DDQN 则智能地选择原地等待，日均等待时长达到 2.37 h，避免出租车盲目空驶，减少空驶里程。

2.3.3 参数敏感性分析

图 9 进一步测试车辆电池容量、充电速率和电动出租车总体数量的影响，测试 DDQN 模型的稳定性。

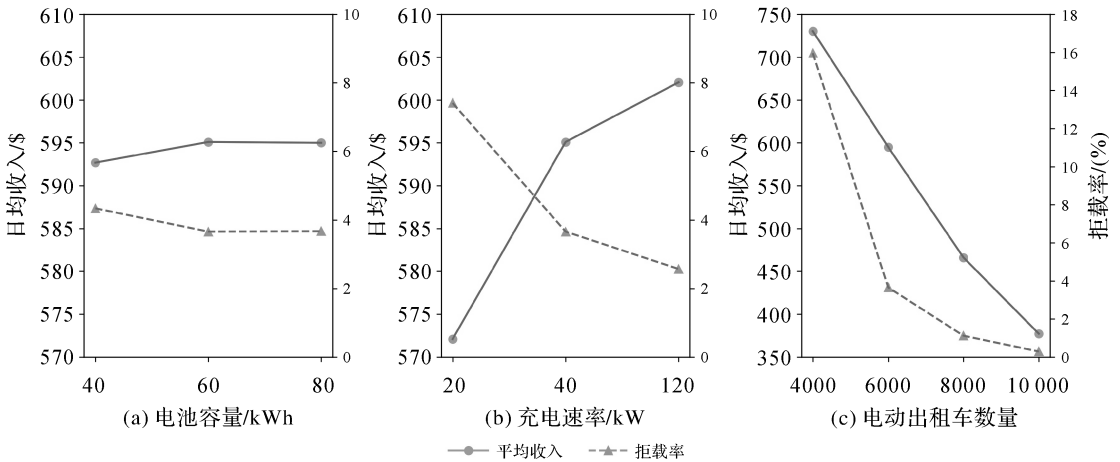


图 9 参数敏感性分析

Fig.9 Parameter sensitivity analysis

电动汽车的电池容量与其续航里程线性正相关。随着电动容量增加,电动出租车可以满足更多的出行需求,日均收入小幅度增加。当电池容量从 60 kWh 进一步提升至 80 kWh 时,出租车运营收益并未得到明显增加。

充电速率越高,充电时间越短。随着充电速率的提高,电动出租车运营不断优化,出行需求拒载率逐步降低。如图 9(b)所示,从 20 kW 慢充提

升至 40 kW 的快充时出租车运营表现提升较大。当充电速率进一步提高至 120 kW 时,电动出租车能够满足 97.5% 的出行需求,获得最高 602 \$ 的日均收入,达到最佳运营表现。建设更多的高速充电站对推广电动出租车意义重大。

图 9(c)给出了不同数量的电动出租车的运营结果。随着车辆数量从 4000 辆增加到 10000 辆,电动出租车服务更加充分,出行需求的总拒载率



从 16% 下降到 0.3%; 但是, 更多的出租车造成竞争加剧, 电动出租车的平均收入出现下降, 日均收入逐渐降低。10 000 辆电动出租车的日均收入约为 350.5 \$。

### 3 结 论

针对电动出租车充电时间长、有效运营时间短等问题, 本文基于深度强化学习对出租车运营决策过程进行建模, 构建 DDQN 模型进行复杂的载客、空驶、等待、充电决策, 最大化电动出租车司机未来收益, 减少司机的充电等待时长。试验结果表明, DDQN 高效的充电与空载策略能够有效协调载客、空驶、等待、充电决策, 日均收入较距离优先策略提高收入约 7%, 拒载比例降低 53%, 充电等待时长减少 70%。本文算法适用于自动驾驶出租车、电动货车等, 能够促进电动出租车的广泛应用, 有利于构建环保、节能与智能的交通出行。后续开展的研究将拓展研究区域, 考虑城市出租车出行的时空规律, 进一步提升模型性能。

### 参考文献:

- [1] 施晓清, 李笑诺, 杨建新. 低碳交通电动汽车碳减排潜力及其影响因素分析[J]. 环境科学, 2013, 34(1): 385-394. SHI Xiaoqing, LI Xiaonuo, YANG Jianxin. Research on carbon reduction potential of electric vehicles for low-carbon transportation and its influencing factors [J]. Chinese Journal of Environmental Science, 2013, 34(1): 385-394.
- [2] 高云. 巴黎气候变化大会后中国的气候变化应对形势[J]. 气候变化研究进展, 2017, 13(1): 89-94. GAO Yun. China's response to climate change issues after Paris Climate Change Conference [J]. Climate Change Research, 2017, 13(1): 89-94.
- [3] 李德仁. 展望大数据时代的地球空间信息学[J]. 测绘学报, 2016, 45(4): 379-384. DOI: 10.11947/j. AGCS. 2016.20160057. LI Deren. Towards geo-spatial information science in big data era [J]. Acta Geodaetica et Cartographica Sinica, 2016, 45(4): 379-384. DOI: 10.11947/j. AGCS.2016.20160057.
- [4] 吴华意, 黄蕊, 游兰, 等. 出租车轨迹数据挖掘进展[J]. 测绘学报, 2019, 48(11): 1341-1356. DOI: 10.11947/j. AGCS.2019.20190210. WU Huayi, HUANG Rui, YOU Lan, et al. Recent progress in taxi trajectory data mining [J]. Acta Geodaetica et Cartographica Sinica, 2019, 48(11): 1341-1356. DOI: 10.11947/j. AGCS.2019.20190210.
- [5] 中国公路学报编辑部. 中国汽车工程学术研究报告·2017 [J]. 中国公路学报, 2017, 30(6): 1-197.

Editorial Department of China Journal of Highway and Transport. Review on China's automotive engineering research progress: 2017 [J]. China Journal of Highway and Transport, 2017, 30(6): 1-197.

- [6] 李清泉. 从 Geomatics 到 Urban Informatics [J]. 武汉大学学报(信息科学版), 2017, 42(1): 1-6. LI Qingquan. From geomatics to urban informatics [J]. Geomatics and Information Science of Wuhan University, 2017, 42(1): 1-6.
- [7] 李德仁, 李清泉, 杨必胜, 等. 3S 技术与智能交通 [J]. 武汉大学学报(信息科学版), 2008, 33(4): 331-336. LI Deren, LI Qingquan, YANG Bisheng, et al. Techniques of GIS, GPS and RS for the development of intelligent transportation [J]. Geomatics and Information Science of Wuhan University, 2008, 33(4): 331-336.
- [8] 涂伟, 李清泉, 方志祥. 一种大规模车辆路径问题的启发式算法 [J]. 武汉大学学报(信息科学版), 2013, 38(3): 307-310, 338. TU Wei, LI Qingquan, FANG Zhixiang. A heuristic algorithm for large scale vehicle routing problem [J]. Geomatics and Information Science of Wuhan University, 2013, 38(3): 307-310, 338.
- [9] 唐炉亮, 阚子涵, 任畅, 等. 利用 GPS 轨迹的转向级交通拥堵精细分析 [J]. 测绘学报, 2019, 48(1): 75-85. DOI: 10.11947/j. AGCS.2019.20170448. TANG Luliang, KAN Zihan, REN Chang, et al. Fine-grained analysis of traffic congestions at the turning level using GPS traces [J]. Acta Geodaetica et Cartographica Sinica, 2019, 48(1): 75-85. DOI: 10.11947/j. AGCS. 2019.20170448.
- [10] VAZIFEH M M, SANTI P, RESTA G, et al. Addressing the minimum fleet problem in on-demand urban mobility [J]. Nature, 2018, 557(7706): 534-538.
- [11] ICCT. Electric vehicle capitals: Accelerating the global transition to electric drive [C] // Proceedings of International Council on Clean Transportation. [S. l.]: ICCT, 2018.
- [12] TU Wei, LI Qingquan, FANG Zhixiang, et al. Optimizing the locations of electric taxi charging stations: a spatial-temporal demand coverage approach [J]. Transportation Research Part C: Emerging Technologies, 2016, 65: 172-189.
- [13] GAN Lingwen, TOPCU U, LOW S H. Optimal decentralized protocol for electric vehicle charging [J]. IEEE Transactions on Power Systems, 2012, 28(2): 940-951.
- [14] KIM H J, LEE J, PARK G L, et al. An efficient scheduling scheme on charging stations for smart transportation [C] // Proceedings of International Conference on Security-Enriched Urban Computing and Smart Grid. Berlin: Springer, 2010, 78: 274-278.
- [15] TIAN Zhiyong, JUNG T, WANG Yi, et al. Real-time charging station recommendation system for electric-

- vehicle taxis[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(11): 3098-3109.
- [16] TU Wei, MAI Ke, ZHANG Yatao, et al. Real-time route recommendations for E-taxis leveraging GPS trajectories [J]. IEEE Transactions on Industrial Informatics, 2020. DOI: 10.1109/TII.2020.2990206.
- [17] Kaelbling L P, Littman M L, Moore A P. Reinforcement learning: a survey [J]. Journal of Artificial Intelligence Research, 1996, 4(1): 237-285.
- [18] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518(7540): 529-533.
- [19] Kober J, Bagnell J A, Peters J. Reinforcement learning in robotics: a survey [J]. The International Journal of Robotics Research, 2013, 32(11): 1238-1274.
- [20] GAO Yong, JIANG Dan, XU Yan. Optimize taxi driving strategies based on reinforcement learning [J]. International Journal of Geographical Information Science, 2018, 32(8): 1677-1696.
- [21] VERMA T, VARAKANTHAM P, KRAUS S, et al. Augmenting decisions of taxi drivers through reinforcement learning for improving revenues [C] // Proceedings of the Twenty-Seventh International Conference on Automated Planning and Scheduling (ICAPS 2017). [S.l.]: ICAPS, 2017: 409-417.
- [22] 荆朝霞, 郭文骏, 郭子暄. 基于多代理技术的电动出租车运营实时仿真系统及应用 [J]. 电力系统自动化, 2016, 40(7): 83-89.  
JING Zhaoxia, GUO Wenjun, GUO Zixuan. Real-time simulation of electric taxi operation based on multi-agent technology [J]. Automation of Electric Power Systems, 2016, 40(7): 83-89.
- [23] TSENG C M, CHAU S C K, LIU Xue. Improving viability of electric taxis by taxi service strategy optimization: a big data study of New York city [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(3): 817-829.
- [24] 高阳, 陈世福, 陆鑫. 强化学习研究综述 [J]. 自动化学报, 2004, 30(1): 86-100.
- GAO Yang, CHEN Shifu, LU Xin. Research on reinforcement learning technology: a review [J]. Acta Automatica Sinica, 2004, 30(1): 86-100.
- [25] BELLMAN R. Dynamic programming and lagrange multipliers [J]. Proceedings of the National Academy of Sciences of the United States of America, 1956, 42(10): 767-769.
- [26] WATKINS C J C H, DAYAN P. Technical note: Q-learning [J]. Machine Learning, 1992, 8(3-4): 279-292.
- [27] Tesauro G. Temporal difference learning and TD-Gammon [J]. Communications of the ACM, 1995, 38(3): 58-68.
- [28] VAN HASSELT H, Guez A, Silver D. Deep reinforcement learning with double Q-learning [C] // Proceedings of the 30th AAAI Conference on Artificial Intelligence. [S.l.]: AAAI Press, 2016: 302-315.
- [29] LIN L J. Reinforcement learning for robots using neural networks [D]. Schenley Park Pittsburgh, PA, United States: Carnegie Mellon University, 1993.
- [30] BAUER G S, GREENBLATT J B, GERKE B F. Cost, energy, and environmental impact of automated electric taxi fleets in Manhattan [J]. Environmental Science & Technology, 2018, 52(8): 4920-4928.

(责任编辑:丛树平)

收稿日期: 2019-12-16

修回日期: 2020-06-07

第一作者简介: 叶浩宇(1996—),男,硕士生,研究方向为城市时空大数据分析。

First author: YE Haoyu (1996—), male, postgraduate, majors in urban spatial-temporal big data analysis.

E-mail: yehaoyuchn@whu.edu.cn

通信作者: 涂伟

Corresponding author: TU Wei

E-mail: tuwei@szu.edu.cn