计 算 机 研 究 与 发 展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202330881
61(9): 2213−2228, 2024

车辆群智感知中激励驱动的车辆选择与调度方法

王振宁[1,2] 曹 越[1,2] 江 恺[1,2] 林 海[1,2] 周 欢[3]

[1](武汉大学国家网络安全学院 武汉 430072)

[2](空天信息安全与可信计算教育部重点实验室(武汉大学) 武汉 430072)

[3](西北工业大学计算机学院 西安 710129)

(zn.wang115@whu.edu.cn)

Incentive-Driven Vehicle Selection and Scheduling Method in Vehicular Crowdsensing

Wang Zhenning[1,2], Cao Yue[1,2], Jiang Kai[1,2], Lin Hai[1,2], and Zhou Huan[3]

[1](*School of Cyber Science and Engineering*, *Wuhan University*, *Wuhan* 430072)

[2](*Key Laboratory of Aerospace Information Security and Trusted Computing* (*Wuhan University*), *Ministry of Education*, *Wuhan* 430072)

[3](*School of Computer Science*, *Northwestern Polytechnical University*, *Xi'an* 710129)

Abstract Vehicular crowdsensing aims to utilize a large number of on-board sensors and computing resources in intelligent vehicles to collect sensing data in a series of areas. Currently, sensing vehicles are usually divided into opportunistic vehicles and participatory vehicles, according to whether the vehicle trajectory can be changed. Here, opportunistic vehicles are with fixed trajectories and cannot be changed. The trajectories of participatory vehicles can be changed based on actual needs. Therefore, selecting the appropriate opportunistic vehicle to complete the sensing task and planning the trajectory of the participatory vehicle is a challenging research problem. We take the different mobility characteristics of the two types of sensing vehicles as the starting point, and manage the two types of vehicles through the crowdsensing platform (CSP). Then, different problems are solved for opportunistic and participatory vehicles, respectively. Specifically, for opportunistic vehicles, it is necessary to select an appropriate vehicle set to complete the sensing task and minimize the cost of CSP. To solve this problem, we propose a reverse auction-based incentive mechanism to select the minimum cost vehicle set to complete the sensing task. It mainly includes two stages: winning vehicle selection and reward payment. Meanwhile, we verify that the proposed method can ensure individual rationality and authenticity. For participatory vehicles, it is necessary to plan the trajectory of each vehicle through CSP scheduling, perform sensing tasks and minimize the cost of CSP. To solve this problem, we propose a deep reinforcement learning-based method to schedule vehicle trajectories and assign different sensing tasks to vehicles. In addition, while minimizing CSP overhead, we also consider the fairness issue of sensing task. The sensing fairness index is introduced to ensure the balance of completion of sensing tasks in different sub-regions. Finally, extensive evaluations based on real-world datasets show that the proposed method performs well and outperforms other benchmark schemes.

Key words vehicular crowdsensing; opportunistic vehicle sensing; participatory vehicle sensing; fairness; reverse auction; deep reinforcement learning

收稿日期: 2023-10-31; **修回日期**: 2024-04-29

基金项目: 国家重点研发计划-政府间国际科技创新合作项目(2022YFE0139300); 湖北省重点研发计划项目(2023BAB014)

This work was supported by the National Key Research and Development Program of China-Intergovernmental International Science and Technology Innovation Cooperation Project (2022YFE0139300) and the Hubei Provincial Key Research and Development Program (2023BAB014).

通信作者: 曹越 (yue.cao@whu.edu.cn)

摘 要 车辆群智感知旨在利用智能车辆配备的车载传感器和计算资源，收集一系列区域的感知数据. 目前，根据车辆轨迹是否可更改，通常可将感知车辆分为机会型车辆和参与型车辆. 其中，机会型车辆轨迹路线固定，不可随意更改. 而参与型车辆轨迹路线可根据现实需求进行更改. 因此，如何选择合适的机会型车辆完成感知任务，以及如何规划参与型车辆的轨迹是一项挑战性研究问题. 这里，以2种类型感知车辆的不同移动特性为出发点，通过群智感知平台（CSP）管理2种类型的车辆，并分别针对机会型车辆和参与型车辆解决不同的问题. 首先，针对机会型车辆，需选择特定的车辆集合，以完成感知任务并最小化CSP开销. 为解决此问题，提出一项基于反向拍卖的激励机制以选择开销最小的车辆集合完成感知任务，主要包括获胜车辆选择和报酬支付2个阶段，同时验证了所提方法可保证机会型车辆的个体合理性和真实性；其次，针对参与型车辆，需通过CSP调度以规划每个参与型车辆的轨迹，执行感知任务并最小化CSP的开销. 为解决此问题，提出一项基于深度强化学习的方法以调度车辆行驶轨迹，为车辆分配不同的感知任务. 此外，在最小化CSP开销的同时，还考虑感知任务执行的公平性问题，引入感知公平指数以确保不同子区域感知任务完成的均衡性. 最后，基于真实世界数据集的广泛评估表明，所提方法效果良好，并优于其他基准方案.

关键词 车辆群智感知；机会型车辆感知；参与型车辆感知；公平性；反向拍卖；深度强化学习

中图法分类号 TP391

近年来，随着5G/6G通信技术和传感器技术的快速发展，智慧城市已进入智能万物互联时代[1-3]. 目前，移动群智感知（mobile crowdsensing, MCS）已经应用于智慧城市建设[4-5]. 相较于传统传感器网络，MCS具有数据采集成本低、设备维护方便、系统架构可扩展等优点[6]. 此外，泛在移动的车辆在智慧城市中发挥了重要作用. 随着网络化、智能化的发展，智能车辆已从一种便捷的交通工具转变为强大的移动感知、计算和存储平台[7]. 因此，车辆群智感知（vehicular crowdsensing, VCS）为智慧城市的发展带来了更多可能性. 具体来说，车辆作为VCS中的基本感知单元，通过支持车联网服务，可实现城市道路交通场景中人、车、环境的协同[8-9]. 同时，由于车辆具有移动性、操作灵活等特点，可支持智慧城市中大规模的自然和社会感知需求，这对于智慧城市的运作与发展都至关重要[10-11].

在VCS中，根据车辆参与感知任务的方式，可将车辆分为机会型车辆和参与型车辆[12]. 其中，机会型车辆依赖其日常的行驶路径执行感知任务，因此无法改变其行驶路线；相反，参与型车辆可按需改变行驶路线，从而可主动前往指定地点执行感知任务[13]. 图1显示了2种类型车辆的示意图. 其中，机会型车辆不会改变其原始轨迹，仅完成处于其途经既定行驶轨迹上分配的感知任务. 例如，公交车等公共车辆可被视为城市场景中的机会型车辆. 一般来说，公共车辆因具有既定行程路线，其路线不可更改. 然而，由于机会型车辆的固定路线的范围有限，可能导致任务覆盖率不足. 因此，参与型车辆的出现弥补了机会型车辆的部分不足. 这里，作为感知实体的参与型车辆，一般可视为是处于空闲状态的出租车或私家车. 由于此类车辆没有既定行驶轨迹，因此可根据感知任务的位置，主动改变其轨迹来完成任务. 与机会型车辆相比，除通过完成感知任务获得报酬奖励外，参与型车辆也可获得因完成该任务产生的相应绕行报酬奖励. 值得注意的是，参与型车辆可能会因被分配到较远距离的感知任务，从而要求高额报酬奖励，或将不可避免地增加平台运营开销.



Fig. 1 Illustration of opportunistic and participatory vehicles sensing

图1 机会型车辆和参与型车辆感知示意图

通常，执行感知任务需要花费额外的时间或经济成本[14]. 由于大多数车辆具备自私和理性特征，在缺乏奖励报酬的情况下，驱动其执行感知任务往往不可行[15]. 因此，需设计适当的奖励机制，以激励车辆参与感知任务. 这里，激励机制主要从经济学角度

引入，并逐渐应用于群智感知[16]. 目前，常用的方法包括拍卖理论、博弈论、契约论等. 此外，由于参与型车辆没有既定行驶轨迹，其需平台统一决策以确定其轨迹，而深度强化学习方法（deep reinforcement learning, DRL）可通过与环境交互，为参与型车辆提供最优决策，并且可适应大规模复杂场景[17-19].

本文将以挖掘2种类型感知车辆的不同移动特性为出发点，分别研究这2种类型车辆面向城市感知场景中的实际问题，即机会型车辆选择问题和参与型车辆调度问题. 具体来说，群智感知平台（crowd sensing platform, CSP）管理1组机会型车辆和1组参与型车辆，通过选择最优的机会型车辆、调度参与型车辆的行驶轨迹，分别完成感知任务.

本文通过选定矩形目标区域，并将其划分为大小相同的子区域网格，车辆经过该区域网格，即代表完成1次感知任务. 此外，本文通过引入感知公平性指数，以保证感知区域均衡性，即不同子区域网格内感知任务完成次数的均衡性. 为解决所提问题，针对机会型感知车辆，本文应用基于反向拍卖的激励方法，通过选择最优的获胜机会型车辆集合完成感知任务，并设计了相应的报酬支付算法；针对参与型感知车辆，则应用基于深度确定性策略梯度 (deep deterministic policy gradient, DDPG) 的方法指导CSP决策，以确定参与型车辆行驶轨迹，同时完成相应的感知任务. 本文所提2种方法分别应用于2种类型车辆的感知场景，并可与2种类型车辆的移动特性相匹配，其目的都是在完成感知任务的同时最小化CSP开销.

本文的主要贡献包括3个方面：

1）针对车辆群智感知场景，本文引入CSP对机会型车辆和参与型车辆进行统一管理. 根据机会型车辆和参与型车辆的移动特性，本文分别考虑了机会型车辆选择问题和参与型车辆调度问题.

2）为了解决上述问题，本文分别提出基于反向拍卖的机会型车辆选择方法和基于DDPG的参与型车辆调度方法. 这2种方法可与2种类型车辆的移动特性相匹配，并有效完成感知任务.

3）此外，本文考虑感知公平性，以确保不同感知子区域内感知任务完成的均衡性，并在CSP的开销中引入公平性指数. 本文的目标是在2种车辆感知的场景下，都能最小化CSP的开销.

1 相关工作

本节回顾了与车辆群智感知相关的现有工作.

1.1 车辆群智感知

近年来，众多研究已考虑车辆群智感知中的车辆招募和选择问题. Hu等人[20]提出了一项基于可变持续时间的车辆招募方法. 在预算有限的情况下，通过考虑车辆不确定性轨迹确定所选车辆和提供服务时长，后续设计了2阶段招募算法以最大限度利用感知资源. 文献[21]通过考虑声誉和车辆轨迹，提出了一项基于贪心的启发式实时车辆选择算法，通过声誉评估和服务定价设计车辆招募计划，在降低成本的同时最大化感知覆盖范围.

此外，部分研究聚焦车辆路径规划和选择问题. Lai等人[11]研究了包括车辆招募、候选路径分析和路径选择在内的问题. 根据历史车辆轨迹信息分析车辆偏好和兴趣，用于选择最优车辆. 然后，利用最小干扰感知策略提出最大加权路径检测问题，以选择车辆行程中断最少且效益最大的路径. Wang等人[22]通过向车辆推荐多条候选路线（每条路线涵盖不同的感知任务），以驱动车辆最低成本路线的选择. 由于所有车辆共享平台的任务奖励，每辆车的选择策略不可避免地会产生竞争. 因此，将群体选择策略问题建模为多车辆感知博弈模型，通过分析车辆之间的交互，提出了一项分布式路径选择方法以解决此问题.

1.2 基于反向拍卖的激励机制

鉴于车辆的个体合理性和自私性，在没有奖励的情况下，很难主动完成感知任务. 因此，需引入激励机制以促使车辆完成感知任务，例如博弈论、拍卖理论、契约理论等，而反向拍卖则是其中一种可行的方法.

近年来，车辆群智感知研究已应用了反向拍卖方法. Fan等人[23]提出了一项称为Hector的方法，用于车辆感知网络的联合轨迹调度和激励机制. 其中，车辆轨迹被映射到时空约束集合中，以降低问题解决的复杂性. 同时，设计了基于反向拍卖的激励机制，以激励车辆通过绕行实现更广的任务覆盖范围. Gao等人[24]研究了基于车辆非确定性的激励机制，通过建立关联平台和车辆间的反向拍卖模型以激励车辆参与感知任务. 此外，还考虑了车辆的感知数据质量，提出了相应的获胜者选择方法和支付确定方法. Chen等人[25]提出了一项基于时效性的激励机制，建立了服务请求者和车辆间的反向拍卖模型，通过将拍卖问题转化为非单调子模块最大化问题，实现预算约束下的请求者的效用提升.

1.3 深度强化学习

随着车辆群智感知系统的不断发展，需采用新

的技术和方法以解决系统规模和复杂性不断增长带来的挑战. 由于 DRL 方法通过将优化问题建模为马尔可夫决策过程(Markov decision processes, MDP)，并基于智能体与环境交互的不断迭代获得最优决策，因此在解决复杂和动态问题方面展现了独特优势. Zhao 等人[26]研究了非合作车辆感知方法，以动态确定感知任务的单位完成成本，通过车辆社交属性激励相关车辆参与感知任务. 为了保障决策的长期最优性，Xu 等人[27]结合生成对抗网络和 DRL 以解决感知任务分配问题，并根据环境灵活自适应调整动作. 这里，该文作者集成生成对抗网络到 DRL 训练过程，并在训练阶段应用双深度 Q 学习方法，以确保感知决策阶段的最优策略.

随着车辆的智能化发展，近期研究还考虑了分布式感知场景，车辆也可充当决策智能体. Ding 等人[28]提出一项基于图卷积协作的多智能体强化学习方法 GCC-MARL，辅助出租车进行分布式决策和全局目标的协同优化. 其中，每辆车充当代理，通过引入信任分配方式以促进车辆间合作，同时结合图卷积网络收集复杂道路网络的空间特征指导车辆的决策动作. Liu 等人[29]提出一项分布式多智能体 DRL 方法，用于节能和分布式无人车辆调度. 该方法将卷积神经网络提取的特征作为演员-评论家网络的输入，生成实时动作并引导无人车辆的轨迹. 此外，使用长短期记忆网络和分布式优先经验回放以实现更优探索策略，通过该方式降低了无人车辆执行感知任务能耗.

1.4 动 机

当前工作主要从单一车辆类型方面进行考虑，未根据机会型车辆和参与型车辆的移动特性进行比较. 因此，本文则以 2 种类型感知车辆的不同移动特性为出发点，分别研究 2 种类型车辆在城市感知场景中的实际问题，即机会型车辆选择问题和参与型车辆调度问题. 然后针对这 2 种问题，根据车辆的移动特性匹配相应的解决方法. 具体来说，本文首先提出了基于反向拍卖的激励机制解决机会型车辆选择问题；然后提出基于 DDPG 的方法解决参与型车辆调度问题. 此外，根据本文模型特性，首次在车辆群智感知场景引入感知公平性指数以保证不同子区域感知任务完成的均衡性. 本文的目标则是在 2 种类型车辆的感知场景下都能最小化 CSP 的开销.

2 系统模型

本节主要介绍 2 种类型车辆的基本定义，然后给出相应的问题表述，并提出本文的优化目标.

2.1 模型定义

本文考虑 1 组机会型车辆、1 组参与型车辆和 CSP 共同存在的场景. 其中，CSP 负责发布 1 组感知任务，并管理和调度以上 2 种类型的车辆分别完成感知任务. 本文根据 2 种类型车辆的移动特性，分别提出了不同的问题，并选择不同的方法以匹配所提问题. 具体来说，针对机会型车辆，由于其轨迹无法更改，本文提出机会型车辆选择问题，并设计基于反向拍卖的激励机制解决此问题. 针对参与型车辆，由于其轨迹可以更改，本文提出参与型车辆调度问题，并设计基于 DRL 的方法调度车辆轨迹解决此问题.

假设在矩形目标区域内，存在 M 辆机会型车辆，表示为 $U=\{u_1,u_2,\cdots,u_M\}$；存在 N 辆参与型车辆，表示为 $V=\{v_1,v_2,\cdots,v_N\}$，2 种类型的车辆都需通过 CSP 进行调度与决策. 此外，本文将矩形目标区域划分为一组大小相同的子区域网格，用 $\mathcal{L}=\{1,2,\cdots,L\}$ 表示子区域网格的序号.

本文中，感知任务被定义为特定时间段内，于目标区域中收集的特定数据，如图像或视频. 为简化表述，每个子区域内都存在 1 个感知任务，当任一车辆经过当前子区域，则视为完成 1 次感知任务. 对于每个子区域内的感知任务，规定车辆收集不超过 η 次数据，即每个子区域的感知次数不能超过 η 次. 此外，设 CSP 执行感知任务的总预算为 B，即需确保车辆执行感知任务的开销小于 B.

定义整个感知窗口为时间 D，每个感知任务在时间 D 内均有效，车辆只有在时间范围内到达才能完成感知任务. 本文假设时间窗口为 $[0,D]$，同时时间也可离散化为 T 个时隙，表示为 $\{1,2,\cdots,t,\cdots,T\}$. 在机会型感知场景中，平台以离线方式决策，机会型车辆执行任务，因此不考虑时隙划分；在参与型感知场景中，则需划分时隙为每个参与型车辆进行实时决策.

2.2 机会型车辆模型

对于机会型车辆 $u_m\in U$，其候选轨迹表示为 r_m. 为简单起见，假设车辆只要到达该子区域网格，即可完成 1 次感知任务. 本文定义候选轨迹 r_m 经过的网格数为 s_m^h，表示车辆 u_m 的候选轨迹 r_m 可完成的任务数. 本文将车辆 u_m 执行 1 次感知任务的成本定义为 c_m^l，这包括了电池、存储、CPU 等各种开销. 对于机会型车辆 u_m，其被选中执行感知任务而产生的成本可表示为

$$C_m=c_m^l s_m^h. \tag{1}$$

同时，我们定义二元变量 $x_m=\{0,1\}$ 表示机会型车辆 u_m 是否被 CSP 选中执行感知任务，其中 $x_m=1$ 表示机会型车辆 u_m 被选中执行感知任务，反之 $x_m=0$ 则表示未被选中.

2.3 反向拍卖模型

由于机会型车辆具备自私和理性特征，在没有激励情况下难以主动完成感知任务，因此本文引入基于反向拍卖的激励机制. 如图 2 所示，在机会型车辆选择过程中，CSP 基于反向拍卖的激励机制以选择获胜车辆，并支付相应的报酬.

具体工作流程为

1）CSP 向机会型车辆候选集合 U 宣布拍卖开始.

2）CSP 和机会型车辆进入拍卖阶段. 在拍卖阶段，CSP 扮演拍卖人角色，机会型车辆扮演竞拍人角色. 每个机会型车辆 u_m 报告其投标价格 b_m 至 CSP，声称其执行感知任务的单位开销（注意 $b_m \geqslant c_m^l$，这意味着机会型车辆 u_m 的投标价格 b_m 不低于其执行感知任务的实际单位成本 c_m^l）.

3）根据机会型车辆的出价集合，CSP 将决定获胜车辆集合 $U' \in U$，并依据获胜车辆轨迹完成相应的感知任务.

4）当机会型车辆完成感知任务后，CSP 对所有获胜车辆执行付款，表示为 q_m. 如车辆未完成任何感知任务，则无法获得任何报酬.

根据上述工作流程，CSP 通过反向拍卖选择获胜机会型车辆集合，向其支付相应的报酬. 同时，CSP 需将支付的报酬控制在合理预算范围内，以保证系统长期激励的有效性. 此外，本文所提基于反向拍卖的方法还需具备 2 个属性：

1）个体合理性. 每个获胜车辆都存在非负效用，即机会型车辆完成感知任务获得的报酬不低于其执行任务的实际成本 c_m^l.

2）真实性. 对于每个获胜车辆，其真实性表示为只有按真实成本出价才可获得最大利润，不能通过恶意提高出价以增加自身利润. 因此，机会型车辆需将其投标价格设置为其执行任务的真实成本. 后文将证明此定理.

2.4 参与型车辆模型

由于参与型车辆没有既定行驶轨迹，因此可通过 CSP 调度前往目标子区域完成感知任务，本文将采用基于 DDPG 的方法确定参与型车辆的行驶路线. 如图 3 所示，参与型车辆位于不同子区域网格，每个时隙内 CSP 将决定参与型车辆 v_n 的动作，即下一时隙应处于的子网格区域. 这里，本文定义参与型车辆

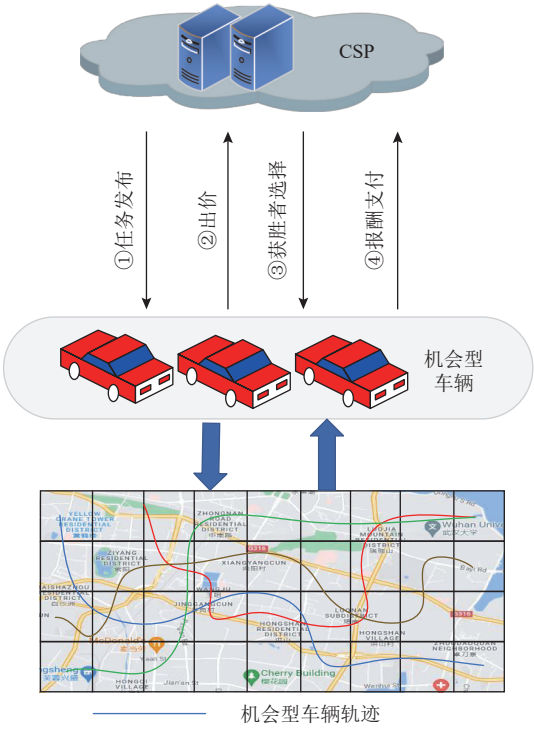


Fig. 2 Illustration of reverse auction-based for opportunistic vehicles selection

图 2 基于反向拍卖的机会型车辆选择图

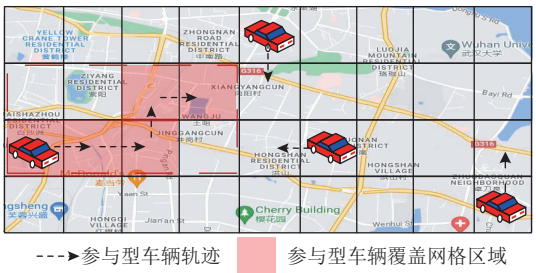


Fig. 3 Illustration of DRL-based for participatory vehicles scheduling

图 3 基于 DRL 的参与型车辆调度示意图

v_n 在时隙 t 的动作为 $a_n^t=\{0,1,2,3,4\}$，表明 CSP 在时隙 t 控制参与型车辆 v_n 从当前子区域水平移动到相邻子区域（1=“前进”，2=“后退”，3=“左转”，4=“右转”），或“停滞”$\left(a_n^t=0\right)$ 在当前子区域内.

对于参与型车辆 v_n，除考虑其固定的执行感知任务的单位成本 c_n^l，还需考虑其绕行所产生的额外距离成本. 给定单位绕行成本为 c_{d}，则参与型车辆所生的距离成本可表示为 $c_{\mathrm{d}}\left(Dis_n-D_n^{\max}\right)$，其中 Dis_n 为车辆 v_n 实际行驶距离，用曼哈顿距离计算. $D_n^{\max}$ 表示车辆 v_n 的计划行驶轨迹距离，如在计划行驶距离内，则 CSP 不需支付距离成本，超过计划距离，CSP 则需支付给参与型车辆 v_n 相应的距离成本. 此外，本文规定参与型车辆 v_n 执行的感知任务数量 $count(v_n)$ 不能超过 g_n.

针对参与型车辆v_n,其执行感知任务的总成本可表示为

$$C_n = c_n^l h_n + c_{\mathrm{d}}\left(Dis_n - D_n^{\max}\right), \tag{2}$$

其中等号右边第1项为参与型车辆执行感知任务的成本,c_n^l为执行任务的单位成本,h_n为车辆当前感知的次数.第2项为执行感知任务所支付的距离成本,c_{d}为单位距离成本.

2.5 问题表述

CSP分别选择1组机会型车辆和调度1组参与型车辆执行感知任务,目标是最小化CSP开销.

首先,定义CSP支付给机会型车辆u_m的报酬为

$$P_m = b_m s_m^h, \tag{3}$$

式(3)中P_m表示机会型车辆u_m执行1次感知任务的出价乘以完成感知任务的次数.

由于机会型车辆需参与反向拍卖过程,CSP向所有机会型车辆支付的报酬为

$$C_{\mathrm{P1}} = \sum_{m=1}^{u_M} x_m b_m s_m^h, \tag{4}$$

其中x_m判断该车辆是否被CSP选中为获胜车辆,CSP仅向被选中的获胜车辆支付相应的报酬.

参与型车辆由CSP进行调度,以完成目标子区域的感知任务.根据式(2),可得CSP向所有参与型车辆支付的报酬为

$$C_{\mathrm{P2}} = \sum_{n=1}^{v_N} C_n. \tag{5}$$

此外,本文通过引入感知公平性定义,以确保不同子区域的感知公平性.由于目标区域被划分为多个子区域网格,且每个子区域网格内可执行多次感知任务,因此需考虑感知子区域的公平性.这里,CSP需保证各子区域网格之间的感知任务完成次数的平等性,避免出现部分子区域感知密度不均问题.因此,本文提出感知公平指数$Fair_Index(U_L)$以反映感知公平性.

基于Jain公平指数[30-31],根据机会型车辆完成感知任务后的状态,本文定义$count_l^u$为所有机会型车辆在子区域网格l内的感知总次数.因此,针对机会型车辆,定义感知公平指数为

$$Fair_Index(U_L) = \frac{\left(\sum_{l=1}^{L} count_l^u\right)^2}{|L|\sum_{l=1}^{L}(count_l^u)^2}, \tag{6}$$

可知$Fair_Index(U_L)\in[0,1]$,当$Fair_Index(U_L)$越接近于1时,代表公平性越好.

同样,针对参与型车辆,定义$count_l^v$为所有参与型车辆在子区域网格l内的感知总次数.因此,可定义参与型车辆的感知公平指数为

$$Fair_Index(V_L) = \frac{\left(\sum_{l=1}^{L} count_l^v\right)^2}{|L|\sum_{l=1}^{L}(count_l^v)^2}. \tag{7}$$

综上,结合感知公平指数,本文重新定义CSP的开销.因此,针对机会型车辆,CSP的开销定义为

$$C_{\mathrm{P}}(U) = \frac{C_{\mathrm{P1}}}{Fair_Index(U_L)}. \tag{8}$$

针对参与型车辆,CSP的开销定义为

$$C_{\mathrm{P}}(V) = \frac{C_{\mathrm{P2}}}{Fair_Index(V_L)}. \tag{9}$$

为降低自身的开销,CSP会选择成本低的机会型车辆参与感知任务,同时尽可能保证感知覆盖的公平性.

2.6 优化问题定义

本文优化目标是在保证感知公平性的前提下,提升感知覆盖范围以最小化CSP开销.因此,针对机会型车辆选择问题,给定优化目标为

$$\mathrm{Q1}: \min C_{\mathrm{P}}(U), \tag{10-1}$$

$$\text{s.t. } x_m = \{0,1\}, \tag{10-2}$$

$$count_l^u < \eta, \tag{10-3}$$

$$C_{\mathrm{P1}} < B, \tag{10-4}$$

其中式(10-2)表示机会型车辆是否被选中的二元变量,式(10-3)表示每个子区域网格的感知次数限制,式(10-4)表示机会型车辆的成本限制,即不能超过CSP的预算B.

针对参与型车辆调度问题,给定优化目标为

$$\mathrm{Q2}: \min C_{\mathrm{P}}(V), \tag{11-1}$$

$$\text{s.t. } count(v_n) < g_n, \tag{11-2}$$

$$count_l^v < \eta, \tag{11-3}$$

$$C_{\mathrm{P2}} < B, \tag{11-4}$$

其中式(11-2)表示参与型车辆v_n执行的感知任务数量不能超过g_n,式(11-3)表示每个子区域网格的感知次数限制,式(11-4)表示支付给参与型车辆的报酬限制,即不能超过CSP的预算B.

由于本文所提出的2个优化问题都为NP-Hard问题,从而可将该优化问题的特殊情况转化为对应的最小加权集覆盖(MWSC)问题来证明,此处省略

具体证明过程.

3 机会型车辆选择问题求解

针对机会型车辆的选择问题，本节提出基于反向拍卖的激励机制，包括获胜车辆选择和报酬支付2个步骤. 此外，本节还证明了该方法保证了机会型车辆的个体合理性和真实性.

3.1 获胜车辆选择

在反向拍卖过程中，机会型车辆作为投标人，CSP作为竞拍人. 由于选择获胜车辆执行感知任务为NP-Hard，本文提出一种近似方法为CSP选择获胜车辆集合. 本文所提机会型车辆获胜选择算法(winner selection algorithm, WSA)主要依靠贪婪思想选取使CSP开销最低的$|M'|$辆车的集合U'，具体过程如下所述.

首先，初始化机会型车辆获胜集合. 通过式(1)计算候选机会型车辆集合中所有车辆的任务完成成本C_m，按递增顺序排序并将新候选车辆集合表示为$U^N=\{u_1,u_2,\cdots,u_m,\cdots,u_{M'},u_{M'+1},\cdots,u_{m'},\cdots,u_M\}$. 然后，在新候选车辆集合$U^N$中选择前$|M'|$辆为初始获胜车辆集合$U'_0=\{u_1,u_2,\cdots,u_m,\cdots,u_{M'}\}$，并计算初始的CSP开销$C_{\mathrm{P}}(U_0)$.

接下来，进入迭代更新过程，此处共包含2层循环过程. 在外层循环中，从车辆$u_{M'+1}$开始至车辆u_M，即针对$\{u_{M'+1},u_{M'},\cdots,u_{m'},\cdots,u_M\}$中的每一辆车有如下内循环过程：依次从后向前替代集合U'_0中的每一辆车，即$\{u_{M'},u_{M'-1},\cdots,u_m,\cdots,u_1\}$. 具体地，在内层循环中，以车辆$u_{M'+1}$为例，用车辆$u_{M'+1}$替代集合$U'$中的每一辆车，并根据每辆车的出价，通过式(8)计算车辆$u_{M'+1}$替换对应车辆u_m后CSP相应的开销$C_{\mathrm{P}}(U_{m'})$.

此时，我们定义

$$\Delta U_{m'}=C_{\mathrm{P}}(U_{m'})-C_{\mathrm{P}}(U_0), \tag{12}$$

式(12)表示车辆$u_{M'+1}$通过替代初始获胜车辆集合U'_0中车辆u_m带来的CSP开销变化. 针对车辆$u_{M'+1}$，当其将集合$\{u_{M'},u_{M'-1},\cdots,u_m,\cdots,u_1\}$中的车辆替换1个轮次后，产生集合$\Delta U(M)=\{\Delta U_1,\Delta U_2,\cdots,\Delta U_{M'}\}$，若存在任意一个$\Delta U<0$，则说明新加入的车辆与原车辆相比降低了CSP的开销，故此时用车辆$u_{M'+1}$代替集合$\Delta U(M)$中对应数值最小的车辆. 上述过程即为一个轮次的内循环过程.

然后，对车辆$u_{M'+2}$~u_M，重复上述内循环过程. 待内循环结束，最终保留在集合U'内的$|M'|$车辆集合即为使CSP开销最小的获胜车辆集合.

注意，在上述过程中，已知子区域网格感知次数和预算受限制，若超出限制，则停止迭代，并将当前车辆删除或终止流程. 具体算法步骤如算法1所示.

算法1. 机会型车辆获胜者选择算法(WSA).

输入：机会型车辆候选集合U_M，预算B，车辆u_m的任务执行成本c_l，车辆u_m的候选轨迹可完成的感知任务数量s_m^h；

输出：获胜者车辆集合U'.

① for each $u_m\in U_M$ do

② 计算每辆车的完成感知任务的成本C_m；

③ end for

④ 根据成本C_m从小到大依次对候选车辆进行排序，得到初始化候选车辆集合$U^N=\{u_1,u_2,\cdots,u_{M'},u_{M'+1},\cdots,u_M\}$；

⑤ 在新候选集合中，选取前$|M'|$辆车作为初始获胜者集合U'_0，并根据式(8)计算初始的CSP开销$C_{\mathrm{P}}(U_0)$；

⑥ for each $u_{m'}\in\{u_{M'+1},u_{M'},\cdots,u_M\}$do

⑦ for each $u_m\in\{u_{M'},u_{M'-1},\cdots,u_1\}$ do

⑧ 依次替代初始候选车辆集合U^N中前$|u_{M'}|$辆车，即$\{u_{M'},u_{M'-1},\cdots,u_1\}$；

⑨ 计算$u_{m'}$代替u_m后对应的$C_{\mathrm{P}}(U_{m'})$，并得到$\Delta U_{m'}=C_{\mathrm{P}}(U_{m'})-C_{\mathrm{P}}(U_0)$；

⑩ 获取集合$\Delta U(M)=\{\Delta U_1,\Delta U_2,\cdots,\Delta U_{m'}\}$；

⑪ if min $\Delta U(M)<0$

⑫ $u_m\leftarrow\arg\min_{u_{m'}}\Delta U(M)$；

⑬ end if

⑭ end for

⑮ end for

⑯ return U'.

3.2 报酬支付

经获胜车辆选择阶段后，CSP需为获胜的机会型车辆支付相应的报酬. 由于机会型车辆完成感知任务会消耗相应的成本，但CSP在选择机会型车辆时已知其期望报酬. 鉴于机会型车辆的自私和理性特征，一般假设其希望获得更高回报. 为此，必须制定统一的支付规则以保障支付合理性. 这里，参考标准VCG(Vickrey-Clarke-Groves)方案的支付规则，本文所提出的支付规则可鼓励机会型车辆参与感知任务，同时确保车辆的个体合理性和真实性.

标准VCG方案中，投标人提交报价并报告其对任务的估价，而不知其他投标人的出价，为此每个获胜者将支付给其他参与者产生的“机会成本”. 这里，

定义投标人S的“机会成本”为在没有投标人S参与情况下的其他投标人的总出价，减去所有其他中标人的出价总和.

本文所提报酬支付算法(payment algorithm, PA)则基于上述方案的思想进行设计. 对于获胜车辆集合U'，首先定义$C_{\mathrm{P}}^{-m}(U)$表示在没有机会型车辆u_m参与情况下CSP的最小开销. 这里，基于算法1可获得在没有机会型车辆u_m参与情况下的最优获胜车辆集合，从而计算此时CSP的最优开销为

$$C_{\mathrm{P}}^{-m}(U)=\frac{C_{\mathrm{P1}}^{-m}}{Fair_Index(U_L)^{-m}}. \tag{13}$$

然后，本文可以得到

$$p_m=C_{\mathrm{P}}^{-m}(U)-C_{\mathrm{P}}(U)>0, \tag{14}$$

其中p_m代表车辆u_m参与拍卖过程中为CSP带来的开销降低的幅度.

根据机会型车辆u_m的出价b_m和对应的感知任务完成数量s_m^h，可得出机会型车辆u_m的效率因子ρ_m，ρ_m根据为CSP带来的开销降低幅度与其需要CSP支付报酬的比值

$$\rho_m=\frac{p_m}{b_m s_h^m}. \tag{15}$$

因此，可确定CSP向获胜车辆u_m支付的报酬表示为

$$q_m=(1+\rho_m)b_m s_h^m. \tag{16}$$

从式(16)可以看出，对于获胜车辆u_m，CSP将支付大于其成本的报酬，从而提高了机会型车辆参与感知任务的积极性. 对于未被选中的车辆，则不向其支付任何报酬. 此外，本文还需要确保CSP的开销不能超过设置预算. 具体算法流程如算法2所示.

算法2. 机会型车辆报酬支付算法(PA).

输入：机会型车辆候选集合U_M，机会型车辆获胜者集合U'，预算B，车辆u_m的任务执行成本c_l和报价b_m，车辆u_m轨迹可完成的感知任务数s_m^h；

输出：向获胜者车辆支付的报酬集合$Q(m)$.

① for each $u_m\in U'$ do

② 根据算法1计算在没有获车辆u_m参与的情况下CSP的获胜车辆集合U^{-m}；

③ 计算当前状态下CSP的最优开销$C_{\mathrm{P}}^{-m}(U)$；

④ 计算$p_m=C_{\mathrm{P}}^{-m}(U)-C_{\mathrm{P}}(U)$；

⑤ 计算$\rho_m=\dfrac{p_m}{b_m s_h^m}$；

⑥ 计算$q_m=(1+\rho_m)b_m s_h^m$；

⑦ $B^*\leftarrow B-q_m$；

⑧ if $B^*<0$

⑨ break；

⑩ end if

⑪ end for

⑫ return $Q(m)=\{q_1,q_2,\cdots,q_m\}$.

3.3 理论分析

本节证明所提基于反向拍卖的机会型车辆选择方法需满足2个重要属性：个体合理性和真实性. 作为车辆帮助CSP完成感知任务的主要动机，个体合理性属性需保证每个机会型车辆都能获得非负奖励. 此外，真实性属性需保证每个机会型车辆无法从不真实的出价中获得更高的报酬.

定理1. 所提出基于反向拍卖的机会型车辆选择方法符合真实性，即每个机会型车辆需要将出价设置为其成本($b_m=c_m^l$).

证明. 首先，本文假设获胜车辆集合中的某机会型车辆u_m报告不真实出价为b'_m ($b'_m>c_m^l$). 然后，根据式(1)可得出当前不真实成本C'_n，由于存在$C'_n>C_n$，进而可根据式(4)得出$C'_{\mathrm{P1}}>C_{\mathrm{P1}}$. 通过计算当前CSP开销$C'_{\mathrm{P}}(U)$，当车辆$u_m$报告不真实出价后，CSP的开销差值可表示为

$$\hat{C}_{\mathrm{P}}(U)=C'_{\mathrm{P}}(U)-C_{\mathrm{P}}(U)=\frac{C'_{\mathrm{P1}}-C_{\mathrm{P1}}}{Fair_Index(U_L)}>0. \tag{17}$$

由(17)可以看出，车辆u_m的不真实出价会增加CSP开销.

然后，根据式(14)，可得出车辆在不真实出价情况下有$p'_m-p_m<0$，进而得到式(15)中有$\rho'_m<\rho_m$. 最后，根据式(16)可得出：

$$\hat{q}_m=q'_m-q_m=\left[1+\left(\rho'_m-\rho_m\right)\right]b_m s_h^m<0. \tag{18}$$

综上，可得知车辆u_m通过不真实出价并不会为自身带来报酬提升. 因此，表明车辆u_m无法通过不真实的投标价格提升自身收益，保证了真实性. 证毕.

定理2. 式(16)中定义的支付规则符合个体合理性，即$\forall u_m\in U', q_m-c_m\geqslant 0$.

证明. 根据本文所提支付规则，对于获胜者集合中的任意车辆u_m，式(14)恒成立，这意味着式(15)中的$\rho_m>0$恒成立. 因此，根据式(16)可得出$q_m>c_m$同样恒成立，从而表明每个获胜车辆都将获得大于自身成本的报酬支付，因此个体合理性得到满足. 证毕.

4 参与型车辆调度问题求解

本节针对参与型车辆的调度问题，提出基于DDPG的方法求解. 本节首先将优化问题近似为MDP，以最小化CSP开销；然后定义MDP中的状态、

动作和奖励函数；最后提出一种基于 DDPG 的方法以解决该问题.

4.1 状态、动作和奖励函数定义

参与型车辆执行感知任务时，由于其行驶轨迹的不确定性，因此可由 CSP 作为智能体，对所有参与型车辆进行统一决策和控制. 在此过程中，由于需即时为车辆做出动作决策，因此本文将时间窗口 $[0,D]$ 离散化为 T 个时隙，并表示为 $\{1,2,\cdots,t,\cdots,T\}$. 这里，本文将介绍基于 DDPG 的参与型车辆调度问题的相关状态空间、动作空间和奖励函数.

1）状态空间（S）. 状态用于表示当前时隙面临的网络环境状况. 相应地，时隙 t 的状态由 $s_t=\{S_1,S_2,B_{\mathrm{r}}^t\}$ 表示. 其中，$S_1=\{loc_{v_1}^t,loc_{v_2}^t,\cdots,loc_{v_n}^t\}$ 表示每个参与型车辆 v_n 在时隙 t 的位置索引（即所处子区域网格的位置索引），$S_2=\{count_1^t,count_2^t,\cdots,count_l^t\}$ 表示在时隙 t 时每个子区域网格的感知次数，B_{r}^t 则表示时隙 t 开始时 CSP 剩余的预算.

2）动作空间（A）. CSP 负责确定每个参与型车辆 v_n 的动作策略，即负责调度参与型车辆轨迹以分配相应的感知任务. 时隙 t 的动作由 $a_t=\{a_1^t,a_2^t,\cdots,a_n^t,\cdots,a_N^t\}$ 表示. 其中，$a_n^t=\{0,1,2,3,4\}$，表明 CSP 在时隙 t 控制参与型车辆 v_n 从当前子区域水平移动到相邻子区域（1=“前进”，2=“后退”，3=“左转”，4=“右转”），或“停滞”$(a_n^t=0)$ 在当前子区域内.

3）奖励函数（R）. 本文目标为最小化 CSP 开销，这与 DRL 最大累积折扣奖励的目标相反. 因此，需设置目标函数的值与奖励的值负相关. 为了最小化 CSP 的长期开销，即时奖励 r_t 被定义为

$$r_t=-C_{\mathrm{P}}^t(V). \tag{19}$$

4.2 从强化学习到深度强化学习

MDP 可用于描述较多决策控制的问题，通常可使用动态规划或线性规划方法解决. 然而，当面对时变且未知的状态和决策时，基于强化学习的方法更适用. 这是因为在强化学习中，智能体在给定的环境中主动学习，并尝试获得更优动作以获得最大累积奖励[32].

这里，Q学习是一种典型的强化学习方法，可描述为在 MDP 中获取最优策略的学习行为，它使用Q表进行值迭代. 其中，智能体需计算每个状态-动作对的Q函数结果，并在每次与环境交互后更新和维护Q表中的Q值.

由于需要维护Q表，Q学习只能应用于小规模动作空间和状态空间的问题，相较于 DRL 方法可在不需先验信息的情况下通过神经网络自动提取高维数据特征，因此在解决大规模状态空间问题时具有优势. 综上，本文提出基于 DDPG 的方法可以解决参与型车辆调度问题.

4.3 基于 DDPG 的参与型车辆调度方法

该方法主要由 2 部分组成，即作为Actor的策略网络和作为Critic的价值网络. 如图 4 所示，首先通过给定初始状态，智能体基于初始动作产生新的状态和奖励，并更新Critic网络；然后，智能体需根据Critic的建议迭代更新Actor网络，在训练出满意的Actor网络前，该迭代会一直持续到最大迭代次数. 对于每次迭代中网络参数的更新，一方面，Critic网络与环境交互，更新自身Q网络奖励的参数 w；另一方面，Critic网络需估计当前状态和动作的Q值以指导Actor网络的更新[33].

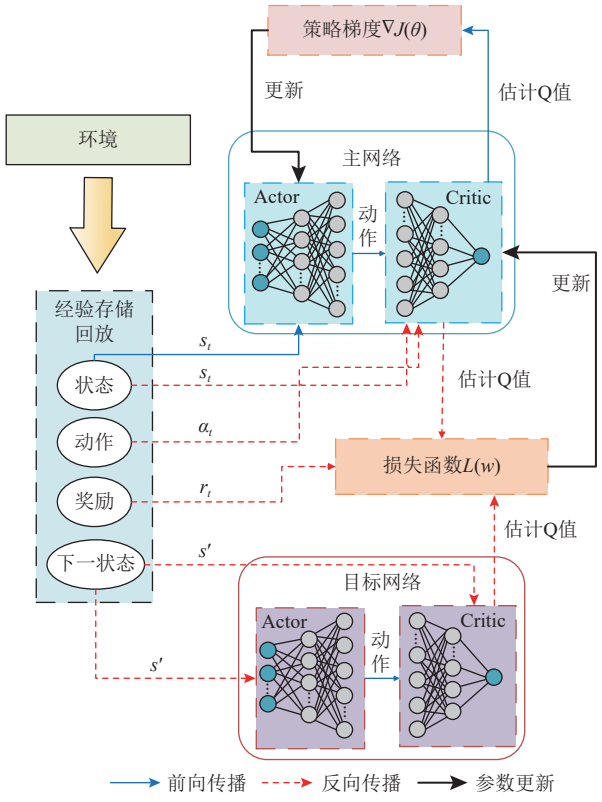


Fig. 4 Illustration of DDPG framework

图 4 DDPG 结构图

本文提出了一种基于 DDPG 的参与型车辆调度算法，以解决所提参与型车辆调度问题. 首先，CSP 作为智能体从环境中收集信息以生成初始状态；然后，主Actor网络根据当前状态 s_t、当前策略 μ 以及随机噪声序列 N 选择动作 a_t：

$$a_t=\mu(s_t,\theta)+N. \tag{20}$$

当获得动作a_t后, 智能体可在每个时隙t将其样本传输元组(s_t,a_t,r_t,s')存储到经验回放池中. 同时, 基于当前状态和动作, 主Critic网络可更新价值网络参数w并计算当前Q值. 本文根据贝尔曼方程将Q值表示为

$$Q^{\mu}(s_t,a_t)=E\left[r(s_t,a_t)+\gamma Q^{\mu}(s',\mu(s',w))\right]. \tag{21}$$

然后, Critic网络通过构造卷积神经网络以模拟贝尔曼方程求解递归Q函数, 并从经验回放池中随机选择一小批样本传输元组更新Critic网络. 这里, 目标Q值由目标Critic网络计算为

$$y_j=r_t+\gamma Q'(s',\mu'(\theta'),w'). \tag{22}$$

为使学习过程稳定并提高算法收敛速度, 智能体从经验回放池中随机选择小批量样本传输元组(s_t,a_t,r_t,s')以消除耦合时间序列训练数据. 在每次迭代中, 本文通过计算y_j并将其发送到主Critic网络以将目标Actor网络与目标Critic网络结合, 并最小化损失函数$L(w)$, 表示为

$$L(w)=\frac{1}{m}\sum_{j=1}^{m}(y_j-Q(s_j,a_j,w))^2, \tag{23}$$

其中m表示样本传输元组的数量.

通过使用价值网络参数w和来自经验回放池的样本元组, 主Actor网络使用策略梯度以更新当前策略, 可表示为

$$\nabla J(\theta)=\frac{1}{m}\sum_{j=1}^{m}[\nabla Q(s_j,a_j,w)|_{s=s_j,a=\mu(s_j)}\nabla_{\theta}\mu(s,\theta)|_{(s=s_j)}]. \tag{24}$$

此外, DDPG 使用软更新方法来更新部分参数:

$$w'\leftarrow\tau w+(1-\tau)w', \tag{25}$$

$$\theta'\leftarrow\tau\theta+(1-\tau)\theta', \tag{26}$$

其中τ表示更新参数.

基于 DDPG 算法的详细内容如算法 3 所示. 基于DDPG 的算法将迭代运行, 直至给定轮次设置(行②). 在每一个 episode 中, 模拟环境被初始化, 初始状态则随机生成(行③④). 首先, 在每个时隙中, 主Actor网络会根据当前状态s_t和策略μ生成相应动作, 并根据得到的动作生成新的状态和奖励; 同时, CSP 作为代理可将其样本传输元组(s_t,a_t,r_t,s')存储到经验回放池中(行⑤~⑧). 然后, 基于当前的状态和动作, 主Critic网络可以更新价值网络参数w并计算当前的Q值(行⑩); 目标Critic网络将当前的Q值发送到主Critic网络, 以最小化损失函数$L(w)$(行⑪). 最后, 根据价值网络参数w和来自经验回放池的样本传输元组, 主Actor网络使用策略梯度来更新当前策略(行⑫), 并在当前时隙结束时, 更新目标网络的参数(行⑬).

算法 3. 基于 DDPG 的参与型车辆调度算法.

输入: 参与型车辆候选集合U_M, 预算B, 初始状态s_0, 测试环境参数$\theta,\theta',w,w',\tau,\gamma,m,N$, 最大迭代次数$T$;

输出: 参与型车辆的轨迹策略.

① 随机初始化参数: $\theta,w,\theta'=\theta,w'=w$, 初始化重放存储器组$D$;
② for each steps of episode do
③ 初始化模拟环境设置;
④ 随机生成并接收一个初始状态;
⑤ for t form 1 to T do
⑥ 在Actor网络中, 根据式(18)获取动作a_t;
⑦ 在系统中执行动作a_t, 获得奖励r_t和下一个状态s_t';
⑧ 将(s_t,a_t,r_t,s')存入经验回放池D;
⑨ 从经验回放池D中随机选择一个小批量的样本传输元组(s_t,a_t,r_t,s');
⑩ 根据式(22)计算当前目标Q值;
⑪ 根据式(23)更新主Critic网络参数w的权重;
⑫ 根据式(24)更新主Actor网络参数θ的权重;
⑬ 根据式(25)(26)更新目标Critic网络和目标Actor网络参数;
⑭ end for
⑮ end for

5 实验分析

本节针对所提的基于反向拍卖的机会型车辆选择方法和基于 DDPG 的参与型车辆调度方法进行性能分析, 并通过真实数据集验证了所提方法的有效性, 并优于基准方案.

5.1 实验设置

针对基于反向拍卖的机会型车辆选择问题, 本文使用基于现实世界的 Roma 数据集[34-35]以验证所提方法的有效性. 该数据集收集了意大利罗马市内 320 辆出租车在 30 天内的轨迹数据. 其中, 数据时间为 2014 年 2 月 1 日至 2014 年 3 月 2 日, 车辆轨迹每隔约 15 s 通过经纬度更新 1 次, 数据格式为〈出租车ID, 日期, 时间, 纬度, 经度〉.

由于该数据集的地理范围较大, 本文使用了原始数据集的一部分, 以确保准确发现车辆运动信息. 首先, 本文选取数据集的 GPS 范围位于(41.87, 12.44)和(41.93, 12.50)之间的矩形区域内. 然后, 将此矩形区域划分为大小相同的子区域网格, 每个子区域网

格存在 1 个感知任务. 根据划分精度不同, 任务的稀疏性也不同, 本文将网格的划分精度区间设置为[0.001,0.01]. 当车辆经过 1 次某子区域网格, 则代表完成 1 次感知任务. 最后, 根据出租车的位置和时间生成车辆轨迹, 并与子区域网格进行匹配, 确定每辆车可完成的感知任务.

本文随机选取数据集中[20,100]辆车的前 1 000 个轨迹点所形成的轨迹进行分析, 并且设定获胜车辆数量为候选车辆数量的一半. 此外, 每辆机会型车辆完成任务的单位成本在[0.3,1.1]之间变化. 为了与所提基于反向拍卖的机会型车辆选择方法(WSA)进行比较, 本文提出了 4 个基准方案:

1)随机(Random). CSP 随机选择目标数量的机会型车辆完成感知任务.

2)基于成本的贪心方法(greedy cost, GC). CSP 仅选择完成感知任务单位成本最低的机会型车辆, 而不考虑其他因素.

3)基于感知任务完成数量的贪心方法(greedy task, GT). CSP 仅选择可完成感知任务数量最多的机会型车辆.

4)不考虑感知公平性的反向拍卖方法[24](no-fair_index, NF). CSP 通过反向拍卖方法选择机会型车辆, 但不考虑感知公平性的影响.

针对基于 DDPG 的参与型车辆调度问题, 本文设置与机会型车辆相同的场景, 即在同样 GPS 范围的矩形区域内进行分析. 然后对矩形目标区域以精度 0.01 进行划分, 将目标区域设置为 6×6 大小的子区域网格, 每个子区域网格最大感知次数设为 10. 对于所提出的基于 DDPG 的参与型车辆调度方法, 本文使用具有 2 个隐藏层的全连接神经网络分别作为 Actor网络和Critic网络, 每层隐藏单元的数量为 128. 最后, 针对网络训练过程, 本文设置经验回放池大小为 5 000, 所选样本传输元组的小批量为 64, 学习率则设为 0.001; 针对网络学习过程, 设置折扣因子γ为 0.9, 软参数τ为 0.005.

为了与所提基于 DDPG 的参与型车辆调度方法进行比较, 本文使用 3 个基准方案:

1)AC(Actor-Critic). AC 构建了一个全方位的智能体, 不仅可直接输出策略, 还可通过价值函数实时评估当前策略的质量. 其中, Actor 使用策略函数生成动作并和环境交互; Critic 使用价值函数评估 Actor 表现, 并指导 Actor 下一阶段动作.

2)随机(Random). CSP 随机做出动作决策, 确定参与型车辆的轨迹路线, 并完成相应的感知任务.

3)贪心方法(Greedy). 针对每辆参与型车辆, CSP 选择相邻子区域网格中感知次数最少的子区域网格作为当前车辆的下一个动作.

5.2 机会型车辆实验结果分析

本节对本文所提出的 WSA 方法的性能进行分析. 图 5 比较了当候选车辆数量变化时的 CSP 开销变化, 其中图 5(a)的网格区域划分精度为 0.005, 图 5(b)的网格区域划分精度为 0.01. 此时, 设置单位任务成本在[0.1,0.3]之间. 可以看出, 随着车辆数量的增加, CSP 的开销逐渐增加, 这是由于有更多机会型车辆参与感知任务, CSP 必然需要支付更多报酬, 所以导致开销增长. 此外, 本文所提出的 WSA 方法具有最低开销, 验证了该方法所选的获胜机会型车辆集合的最优性. Random 效果最差, 主要因为随机选择无法带来任何优化. NF 结果接近于 WSA, 这是因为 NF 也通过获胜者选择方法选取最优车辆集合. 然而, 由于 NF 为不考虑感知公平性下的最优解, 这验证了 WSA 方法在感知公平性上的优越性.

图 6 比较了 CSP 的开销随机会型车辆执行感知

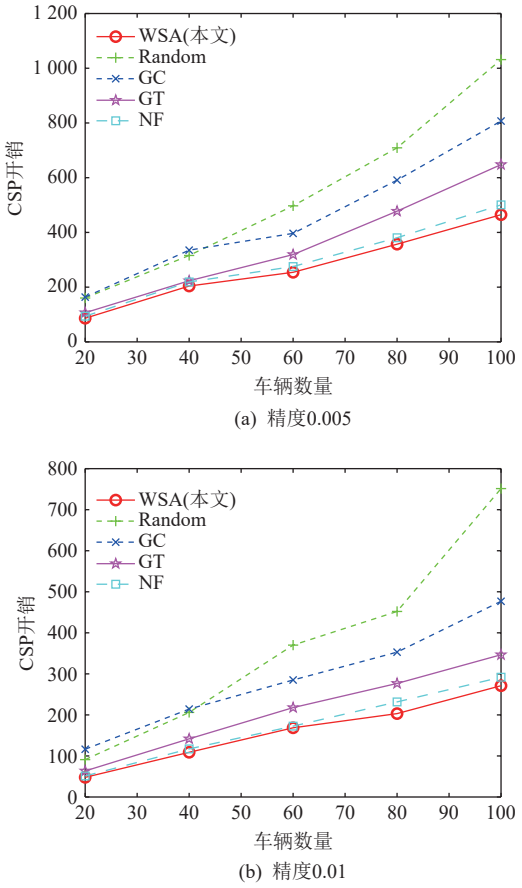


Fig. 5 CSP overhead changes with the number of opportunistic vehicles

图 5 CSP 开销随机会型车辆数量的变化

任务的单位成本的变化情况, 其中图 6(a)的网格区域划分精度为 0.005, 图 6(b)的网格区域划分精度为 0.01. 此时, 将候选车辆集合设定为 60. 可以发现, WSA 略优于 NF 和 GT, 并大幅优于 Random 和 GC. 这是因为 GC 仅选取感知成本最低的机会型车辆, 并没有考虑车辆可完成的任务数量, 因而导致较低的公平性指数, 产生更高开销. 此外, 随着感知任务执行成本的增长, CSP 的开销也随之逐渐增长.

图 7 和图 8 比较了 WSA 与 NF 在感知公平性方面的性能, 此时候选车辆数量设为 100 辆. 其中, 图 7 可视化了在精度 0.01 下 WSA 方法和 NF 方法中各网格感知次数的热力图, 图 8 可视化了在精度 0.005 下 WSA 方法和 NF 方法中各网格感知次数的热力图. 这里, 子网格颜色深浅代表了对应子区域网格感知任务的完成次数. 可以看出, 与 NF 相比, WSA 拥有更好的感知公平性. 此外, 图 7(a)中网格区域完成的最大感知任务次数为 72, 而图 7(b)中网格区域完成的最大感知任务次数为 102. 图 8(a)中网格区域完成的最大感知任务次数为 55, 图 8 (b)中网格区域完成的最大感知任务次数为 59. 这里, 尽管最大感知任务

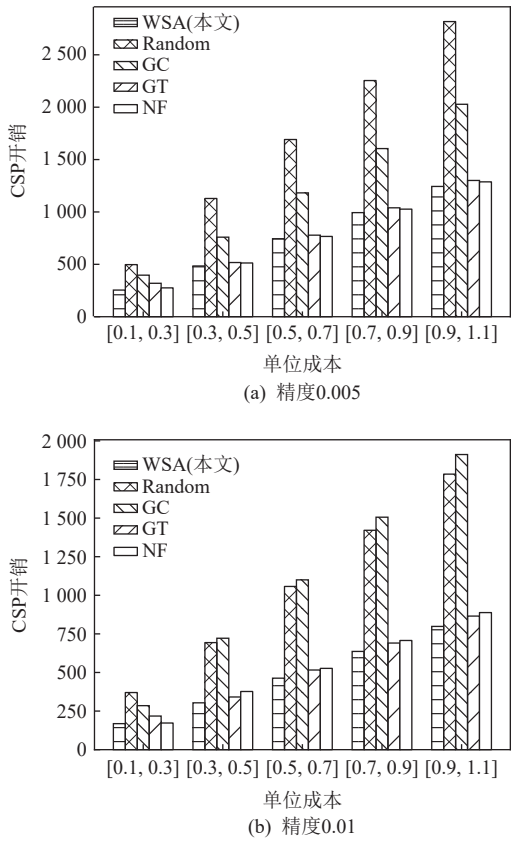


Fig. 6 CSP overhead changes with the unit task cost of opportunistic vehicles

图 6 CSP 开销随机会型车辆单位任务成本的变化

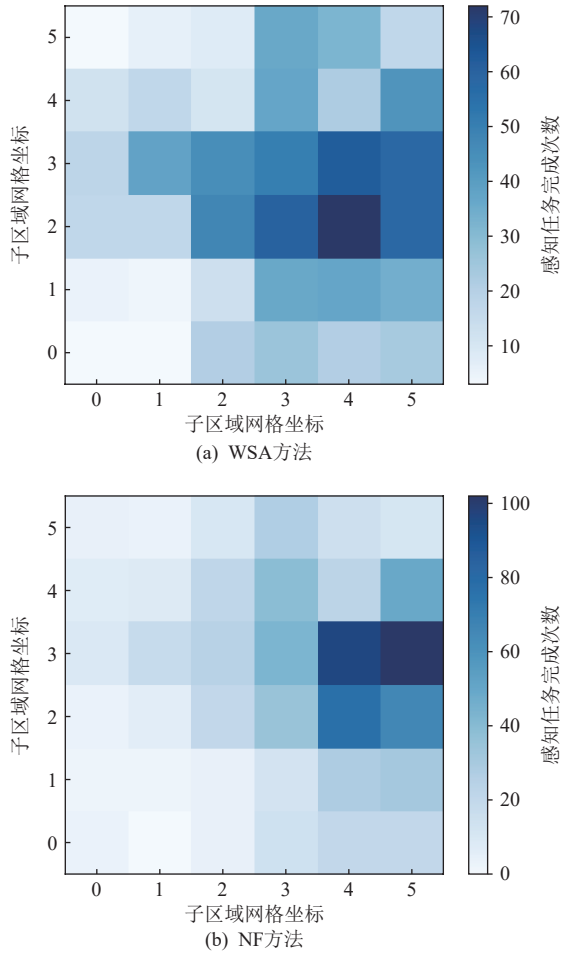


Fig. 7 Visualization of the number of completed sensing tasks in different sub-region grids at 0.01 accuracy

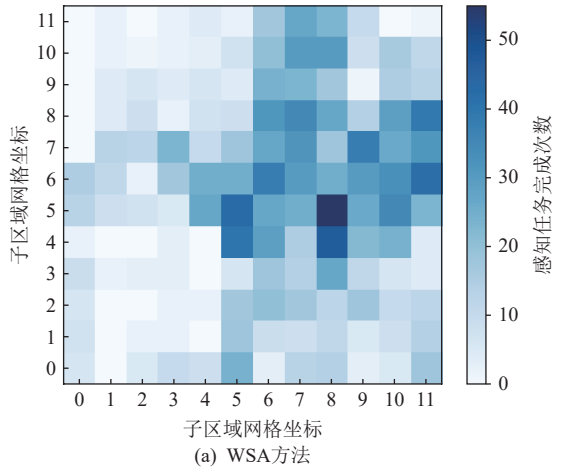
图 7 精度 0.01 时不同子区域网格感知任务完成次数可视化

次数的数值接近, 但是可看出感知公平性差异明显.

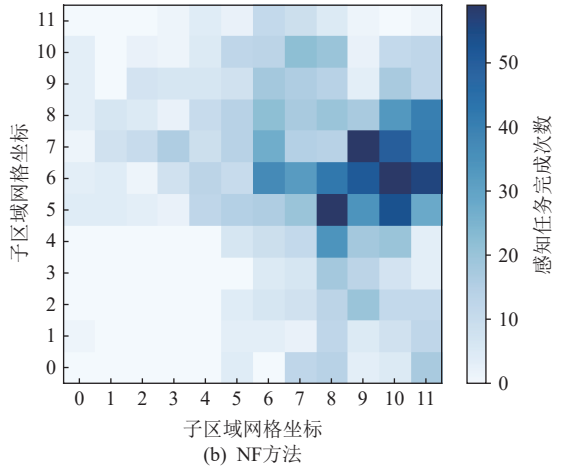
此外, 本文对算法 2 的有效性进行验证, 并选取其中10辆获胜机会型车辆的结果进行分析, 如图 9 和表 1 所示. 在图 9 中, 车辆数据点被表示为(单位成本, 单位报酬), 可看出车辆的数据都在图 9 中的左上方, 这说明每辆获胜的机会型车辆的单位报酬都大于其单位成本, 从而符合预期. 此外, 根据表 1 可以看出, 获胜车辆单位报酬的大小与其对 CSP 开销降低带来的贡献程度有关, 贡献越大, 获得的单位报酬就越高. 相反, 感知任务完成的次数与单位报酬并无直接关系, 这是因为尽管车辆完成较多感知任务, 但对 CSP 公平性并没有带来明显改善, 因此无法获得较高报酬.

5.3 参与型车辆实验结果分析

本节对本文所提出的基于 DDPG 的参与型车辆调度方法进行性能分析. 首先进行收敛性分析, 本文选取6辆车在6×6的网格区域内进行实验, 设置参与



(a) WSA方法



(b) NF方法

Fig. 8 Visualization of the number of completed sensing tasks in different sub-region grids at 0.005 accuracy

图 8 精度 0.005 时不同子区域网格感知任务完成次数可视化

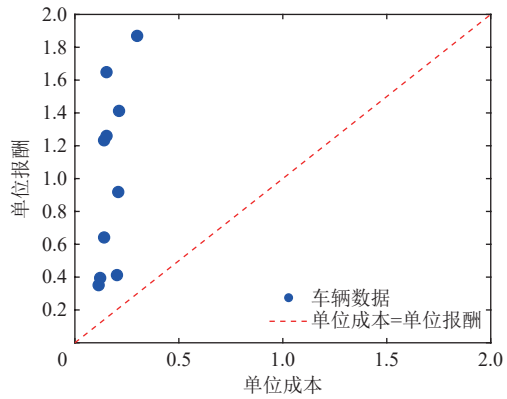


Fig. 9 Validation of PA method effectiveness

图 9 PA 方法有效性验证

型车辆单位任务成本在[0.1,0.3]之间. 图 10 总共绘制了 1 600 个 episode 的累积平均奖励的收敛曲线图. 正如预测的那样, 与 AC 算法相比, 基于 DDPG 的参与型车辆调度方法表现出更快的收敛速度和更好的性能. 可以看出, 随着 episode 训练轮次的增加, 平均奖励迅速增加, 尽管在 200~400 轮次出现波动, 但仍在 1 000 轮次后收敛到相对稳定的值. 此外, 可看出 AC 收敛速度则更加缓慢, 波动也较大, 这是因为 AC 更新过程中没有经验回放池, 导致收敛速度变慢. 而基于 DDPG 的参与型车辆调度方法利用经验回放池和双网络结构, 使神经网络学习更加有效. 因此, 基于 DDPG 的方法比 AC 具有更好的收敛性能. 此外, 图 10 中还可看出, 传统适用于离散动作的 DQN 方法效果表现较差, 收敛速度非常缓慢, 这是由于动作空间过大造成的. 因此本文通过离散动作连续化, 应用 DDPG 方法可以更好地获取参与型车辆动作决策.

Table 1 Related Results of PA Method

表 1 PA 方法相关结果

获胜车辆 ID	单位成本	单位付款	CSP 开销降低的幅度	任务完成次数
370	0.140 326	1.233 298	26.231 32	24
179	0.152 087	1.648 364	35.910 63	24
174	0.208 489	0.918 434	14.908 86	21
290	0.140 510	0.641 551	17.035 37	34
14	0.152 216	1.260 109	37.668 36	34
339	0.202 298	0.412 341	6.721 365	32
19	0.114 158	0.350 992	11.368 04	48
278	0.299 363	1.869 238	31.397 49	20
349	0.212 298	1.412 341	16.721 36	32
30	0.121 529	0.394 767	15.574 60	57

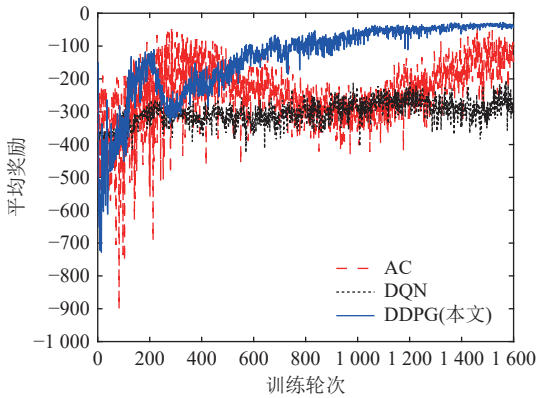
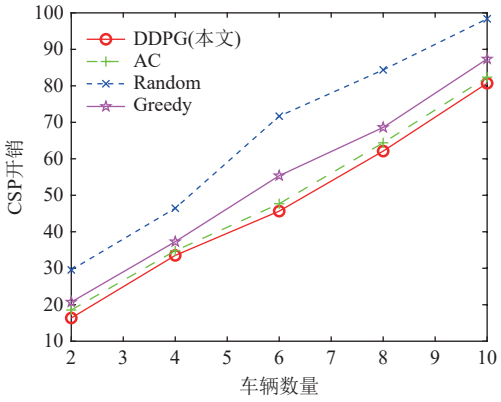


Fig. 10 The average reward trends for each episode

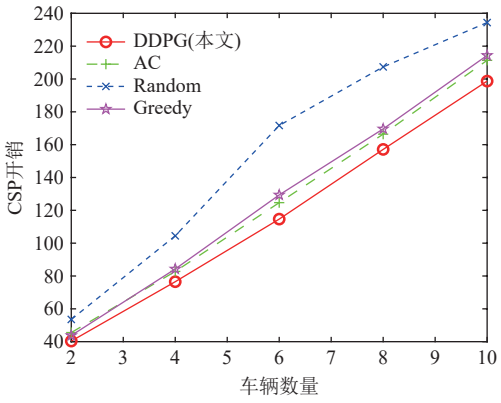
图 10 每 episode 平均奖励变化趋势

其次, 图 11 比较了 CSP 的开销随参与型车辆数量的变化情况(参与型车辆数量设置为 [2,10]), 其中图 11(a)的网格区域划分精度为 0.01, 图 11(b)的网格区域划分精度为 0.005. 可以看出, 随着车辆数量的增加, CSP 的开销逐渐增加, 这是因为有更多参与型车辆参与感知任务, CSP 必然需要支付更多的报酬, 从而导致开销增长. 此外, 本文所提出的基于 DDPG

的方法具有最低开销, 从而验证了该方法的优越性. 相比较, Random 效果最差, 主要因为其无法在感知公平性方面带来任何优化. 尽管 AC 接近于本文所提基于 DDPG 的方法, 但其收敛速度较慢. 此外, Greedy 方法通过贪婪方式获取车辆动作, 容易导致局部最优解, 因此 CSP 开销也较大.



(a) 精度0.01



(b) 精度0.005

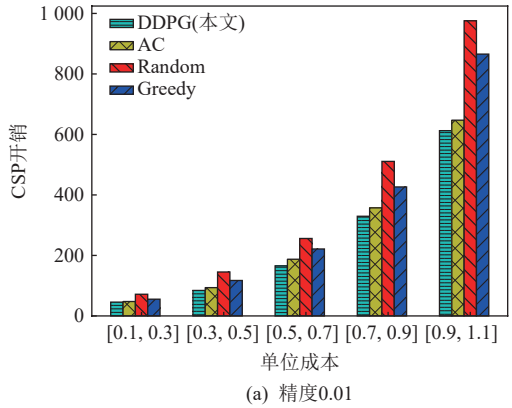
Fig. 11 CSP overhead changes with the number of participatory vehicles

图 11 CSP 开销随参与型车辆数量的变化

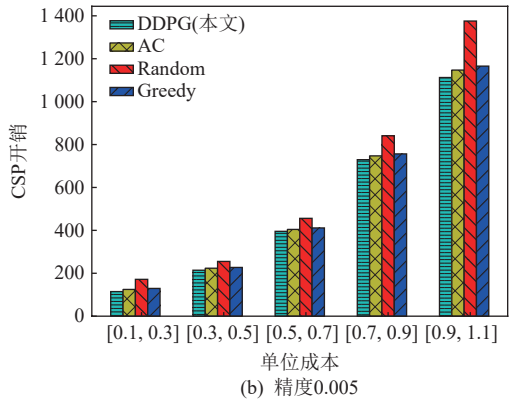
最后, 图 12 比较了 CSP 开销随参与型车辆执行感知任务的单位成本的变化情况(参与型车辆数量设为 6), 其中图 12(a)的网格区域划分精度为 0.01, 图 12(b)的网格区域划分精度为 0.005. 可以发现, 本文所提基于 DDPG 的参与型车辆调度方法略优于 AC, 并大幅优于 Random 和 Greedy, 这种差距随着单位成本的增加而更加显著. 此外, 随着感知任务执行成本的增长, CSP 的开销也随之逐渐增长.

6 总结与展望

本文以挖掘机会型车辆和参与型车辆的不同移动特性为出发点, 分别研究了这 2 种类型的车辆面



(a) 精度0.01



(b) 精度0.005

Fig. 12 CSP overhead changes with the unit task cost of participatory vehicles

图 12 CSP 开销随参与型车辆单位任务成本的变化

向城市感知场景中的实际问题. 针对机会型车辆, 本文应用基于反向拍卖的激励方法, 通过选择最优的获胜机会型车辆集合完成感知任务, 并设计了相应的报酬支付算法. 针对参与型车辆, 应用基于 DDPG 的方法指导 CSP 决策, 以确定参与型车辆行驶轨迹, 并完成相应的感知任务. 此外, 本文引入感知公平性指数以保证不同子区域感知任务完成的均衡性, 保证了在 2 种类型的车辆感知场景下都能最小化 CSP 的开销. 基于真实世界数据集的广泛评估表明, 本文所提方法效果良好, 并优于其他基准方案.

目前, 根据 2 种类型车辆的不同移动特性, 本文选择对应的解决方法以匹配所提问题, 目的是突出 2 种类型车辆的不同移动特性在感知任务完成方面带来的影响. 在未来工作中, 我们计划将这 2 种类型车辆相结合, 并考虑不同类型车辆间在完成感知任务时的相互影响. 据此, 可探索结合 2 种类型车辆的混合感知方法, 以发挥 2 种类型车辆在移动特性方面的优势, 共同完成感知任务.

作者贡献声明: 王振宁提出了实验方案并撰写

论文；曹越提出指导意见并修改论文；江恺对实验方案设计提出意见；林海和周欢参与论文修改.

参 考 文 献

[1] Yu Zhiwen, Ma Huadong, Guo Bin, et al. Crowdsensing 2.0[J]. Communications of the ACM, 2021, 64(11): 76−80

[2] Liu Yutong, Kong Linghe, Chen Guihai. Data-oriented mobile crowdsensing: A comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2019, 21(3): 2849−2885

[3] Capponi A, Fiandrino C, Kantarci B, et al. A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities[J]. IEEE Communications Surveys & Tutorials, 2019, 21(3): 2419−2465

[4] Ma Guoqi, Chen Honglong, Huang Yang, et al. Utility-based heterogeneous user recruitment of multi-task in mobile crowdsensing[J]. IEEE Internet of Things Journal, 2023, 10(11): 9796−9808

[5] Li Xiaoqian, Feng Gang, Liu Yijing, et al. Joint sensing, communication, and computation in mobile crowdsensing enabled edge networks[J]. IEEE Transactions on Wireless Communications, 2022, 22(4): 2818−2832

[6] Ji Guoliang, Yao Zheng, Zhang Baoxian, et al. Quality-driven online task-bundling-based incentive mechanism for mobile crowdsensing[J]. IEEE Transactions on Vehicular Technology, 2022, 71(7): 7876−7889

[7] Liu Luning, Wang Luhan, Lu Zhaoming, et al. Cost-and-quality aware data collection for edge-assisted vehicular crowdsensing[J]. IEEE Transactions on Vehicular Technology, 2022, 71(5): 5371−5386

[8] Tong Yao, Zhu Shuli, Ren Xiaotong, et al. Vehicle inertial tracking via mobile crowdsensing: Experience and enhancement[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1−13

[9] Yang Xinxin, Gu Bo, Zheng Beichen, et al. Toward incentive-compatible vehicular crowdsensing: An edge-assisted hierarchical framework[J]. IEEE Network, 2022, 36(2): 162−167

[10] Ren Yilong, Jiang Han, Feng Xiaoyuan, et al. ACP-based modeling of the parallel vehicular crowd sensing system: Framework, components and an application example[J]. IEEE Transactions on Intelligent Vehicles, 2022, 8(2): 1536−1548

[11] Lai Yongxuan, Xu Yifan, Mai Duojian, et al. Optimized large-scale road sensing through crowdsourced vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(4): 3878−3889

[12] Fu Yanming, Qin Xiaoqiong, Zhang Xian, et al. Hybrid recruitment scheme based on deep learning in vehicular crowdsensing[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(10): 10735−10748

[13] Yu Ruiyun, Wang Pengfei, Bai Zhihong, et al. Participatory sensing: People-centric smart sensing and computing[J]. Journal of Computer Research and Development, 2017, 54(3): 457−473 (in Chinese)
(于瑞云, 王鹏飞, 白志宏, 等. 参与式感知: 以人为中心的智能感知与计算[J]. 计算机研究与发展, 2017, 54(3): 457−473)

[14] Ying Chenhao, Xia Fuyuan, Li Jie, et al. Incentive mechanism based on truth estimation of private data for blockchain-based mobile crowdsensing[J]. Journal of Computer Research and Development, 2022, 59(10): 2212−2232 (in Chinese)
(应臣浩, 夏福源, 李颉, 等. 区块链群智感知中基于隐私数据真值估计的激励机制[J]. 计算机研究与发展, 2022, 59(10): 2212−2232)

[15] Chen Fangzhe, Huang Lianfen, Gao Zhibin, et al. Latency-sensitive task allocation for fog-based vehicular crowdsensing[J]. IEEE Systems Journal, 2022, 17(2): 1909−1917

[16] Xiang Chaocan, Cheng Wenhui, Lin Chi, et al. LSTAloc: A driver-oriented incentive mechanism for mobility-on-demand vehicular crowdsensing market[J]. IEEE Transactions on Mobile Computing, 2023, 23(4): 3106−3122

[17] Li Mengge, Ma Miao, Wang Liang, et al. Multi-task-oriented collaborative crowdsensing based on reinforcement learning and blockchain for intelligent transportation system[J]. IEEE Transactions on Industrial Informatics, 2022, 19(9): 9503−9514

[18] Xu Ying, Wang Yufeng, Ma Jianhua, et al. PSARE: A RL-based online participant selection scheme incorporating area coverage ratio and degree in mobile crowdsensing[J]. IEEE Transactions on Vehicular Technology, 2022, 71(10): 10923−10933

[19] Li Xin, Lei Xinghua, Liu Xiuwen, et al. Collision-free parking recommendation based on multi-agent reinforcement learning in vehicular crowdsensing[J/OL]. Digital Communications and Networks, [2023-05-10].https://doi.org/10.1016/j.dcan.2023.04.005

[20] Hu Miao, Zhong Zhangdui, Niu Yong, et al. Duration-variable participant recruitment for urban crowdsourcing with indeterministic trajectories[J]. IEEE Transactions on Vehicular Technology, 2017, 66(11): 10271−10282

[21] Abdelhamid S, Hassanein H S, Takahara G. Reputation-aware, trajectory-based recruitment of smart vehicles for public sensing[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 19(5): 1387−1400

[22] Wang En, Luan Dongming, Yang Yongjian, et al. Distributed game-theoretical route navigation for vehicular crowdsensing[C/OL]//Proc of the 50th Int Conf on Parallel Processing. New York: ACM, (2021-10-05)[2023-01-15].https://doi.org/10.1145/3472456.3472498

[23] Fan Guiyun, Jin Haiming, Liu Qihong, et al. Joint scheduling and incentive mechanism for spatio-temporal vehicular crowd sensing[J]. IEEE Transactions on Mobile Computing, 2019, 20(4): 1449−1464

[24] Gao Guoju, Xiao Mingjun, Wu Jie, et al. Truthful incentive mechanism for nondeterministic crowdsensing with vehicles[J]. IEEE Transactions on Mobile Computing, 2018, 17(12): 2982−2997

[25] Chen Xianhao, Zhang Lan, Pang Yawei, et al. Timeliness-aware incentive mechanism for vehicular crowdsourcing in smart cities[J]. IEEE Transactions on Mobile Computing, 2021, 21(9): 3373−3387

[26] Zhao Yinuo, Liu C. Social-aware incentive mechanism for vehicular crowdsensing by deep reinforcement learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(4): 2314−2325

[27] Xu Chenghao, Song Wei. Intelligent task allocation for mobile crowdsensing with graph attention network and deep reinforcement learning[J]. IEEE Transactions on Network Science and Engineering, 2023, 10(2): 1032−1048

[28] Ding Rong, Yang Zhaoxing, Wei Yifei, et al. Multi-agent reinforcement learning for urban crowd sensing with for-hire vehicles[C/OL]//Proc of the 40th IEEE Conf on Computer Communications. Piscataway, NJ: IEEE, (2021-07-26)[2023-10-04].https://doi.org/10.1109/INFOCOM42981.2021.9488713

[29] Liu C, Dai Zipeng, Zhao Yinuo, et al. Distributed and energy-efficient mobile crowdsensing with charging stations by deep reinforcement learning[J]. IEEE Transactions on Mobile Computing, 2019, 20(1): 130−146

[30] Ding Lige, Zhao Dong, Cao Mingzhe, et al. When crowdsourcing meets unmanned vehicles: Toward cost-effective collaborative urban sensing via deep reinforcement learning[J]. IEEE Internet of Things Journal, 2021, 8(15): 12150−12162

[31] Liu C, Chen Zheyu, Zhan Yufeng. Energy-efficient distributed mobile crowd sensing: A deep learning approach[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(6): 1262−1276

[32] Zhou Huan, Wang Zhenning, Zheng Hantong, et al. Cost minimization-oriented computation offloading and service caching in mobile cloud-edge computing: An A3C-based approach[J]. IEEE Transactions on Network Science and Engineering, 2023, 10(3): 1326−1338

[33] Zhou Huan, Zhang Zhenyu, Wu Yuan, et al. Energy efficient joint computation offloading and service caching for mobile edge computing: A deep reinforcement learning approach[J]. IEEE Transactions on Green Communications and Networking, 2023, 7(2): 950−961

[34] Cao Xiaofeng, Yang Peng, Lyu Feng, et al. Trajectory penetration characterization for efficient vehicle selection in HD map crowdsourcing[J]. IEEE Internet of Things Journal, 2020, 8(6): 4526−4539

[35] Wang Xiaojie, Ning Zhaolong, Hu Xiping, et al. A city-wide real-time traffic management system: Enabling crowdsensing in social internet of vehicles[J]. IEEE Communications Magazine, 2018, 56(9): 19−25



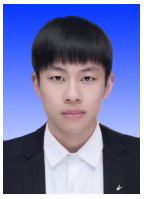
Wang Zhenning, born in 1998. PhD candidate. His main research interests include traffic crowdsensing, incentive mechanisms, and deep reinforcement learning.

王振宁, 1998 年生. 博士研究生. 主要研究方向为交通群智感知、激励机制和深度强化学习.



Cao Yue, born in 1984. PhD, professor. Member of CCF. His main research interests include security protection, network computing, and traffic control.

曹　越, 1984 年生. 博士，教授. CCF 会员. 主要研究方向为安全防护、网络计算、交通控制.



Jiang Kai, born in 1995. PhD candidate. His main research interests include transportation shared service computing, edge intelligence, and federated learning.

江　恺, 1995 年生. 博士研究生. 主要研究方向为交通共享服务计算、边缘智能、联邦学习.



Lin Hai, born in 1976. PhD, associate professor. Member of CCF. His main research interests include cyber security and Internet of things.

林　海, 1976 年生. 博士，副教授. CCF 会员. 主要研究方向为网络安全、物联网.



Zhou Huan, born in 1986. PhD, professor. Senior member of CCF. His main research interests include mobile social network, mobile edge computing, and Internet of vehicles.

周　欢, 1986 年生. 博士，教授. CCF 高级会员. 主要研究方向为移动社交网络、移动边缘计算、车联网.