软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2024,35(3):1485−1501 [doi: 10.13328/j.cnki.jos.006814]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于深度强化学习的 WRSN 动态时空充电调度*

王艺均[1], 冯 勇[1], 刘 明[2], 刘念伯[2]

[1](云南省计算机技术应用重点实验室 (昆明理工大学), 云南 昆明 650500)

[2](电子科技大学 计算机科学与工程学院, 四川 成都 611731)

通信作者: 冯勇, E-mail: fybraver@kust.edu.cn

摘 要: 高效的移动充电调度是构建长生命期、可持续运行的无线可充电传感器网络 (WRSN) 的关键之一. 现有基于强化学习的充电策略只考虑了移动充电调度问题的一个维度, 即移动充电器 (MC) 的路径规划, 而忽略了充电调度问题中的另一维度, 即充电时长调整, 因而仍然存在性能限制. 提出一种基于深度强化学习的 WRSN 动态时空充电调度方法 (SCSD), 建立充电序列调度和充电时长动态调整的深度强化学习模型. 针对移动充电调度中离散的充电序列规划和连续的充电时长调整问题, 使用 DQN 为待充电节点优化充电序列, 并基于 DDPG 计算并动态调整序列中待充电节点的充电时长. 通过分别从空间和时间两个维度的优化, 在避免节点缺电失效的同时, 所提出的 SCSD 可实现充电性能的有效提高. 大量仿真实验结果表明, SCSD 与现有的几种有代表性的充电方案相比, 其充电性能具有明显的优势.

关键词: 无线可充电传感器网络; 深度强化学习; 时空充电策略; 充电序列; 充电时长; 充电性能

中图法分类号: TP393

中文引用格式: 王艺均, 冯勇, 刘明, 刘念伯. 基于深度强化学习的WRSN动态时空充电调度. 软件学报, 2024, 35(3): 1485–1501. http://www.jos.org.cn/1000-9825/6814.htm

英文引用格式: Wang YJ, Feng Y, Liu M, Liu NB. Dynamic Spatiotemporal Charging Scheduling Based on Deep Reinforcement Learning for WRSN. Ruan Jian Xue Bao/Journal of Software, 2024, 35(3): 1485–1501 (in Chinese). http://www.jos.org.cn/1000-9825/6814.htm

Dynamic Spatiotemporal Charging Scheduling Based on Deep Reinforcement Learning for WRSN

WANG Yi-Jun[1], FENG Yong[1], LIU Ming[2], LIU Nian-Bo[2]

[1](Yunnan Key Laboratory of Computer Technology Applications (Kunming University of Science and Technology), Kunming 650500, China)

[2](School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: Efficient mobile charging scheduling is a key technology to build wireless rechargeable sensor networks (WRSN) which have long life cycle and sustainable operation ability. The existing charging methods based on reinforcement learning only consider the spatial dimension of mobile charging scheduling, i.e., the path planning of mobile chargers (MCs), while leaving out the temporal dimension of the problem, i.e., the adjustment of the charging duration, and thus these methods have suffered some performance limitations. This study proposes a dynamic spatiotemporal charging scheduling scheme based on deep reinforcement learning (SCSD) and establishes a deep reinforcement learning model for dynamic adjustment of charging sequence scheduling and charging duration. In view of the discrete charging sequence planning and continuous charging duration adjustment in mobile charging scheduling, the study uses DQN to optimize the charging sequence for nodes to be charged and calculates and dynamically adjusts the charging duration of the nodes. By optimizing the two dimensions of space and time respectively, the SCSD proposed in this study can effectively improve the charging performance while avoiding the power failure of nodes. Simulation experiments show that SCSD has significant performance advantages over several

* 基金项目: 国家自然科学基金 (62062047, 61662042, 61962030)
收稿时间: 2022-04-13; 修改时间: 2022-05-24; 采用时间: 2022-10-11; jos 在线出版时间: 2023-05-17
CNKI 网络首发时间: 2023-05-18

well-known typical charging schemes.

Key words: wireless rechargeable sensor network (WRSN); deep reinforcement learning; spatiotemporal charging scheme; charging sequence; charging duration; charging performance

近年来, 无线充电技术[1,2]的突破性进展为彻底解决无线传感器网络的能量限制问题开辟了新途径. 传感器配备有无线能量接收装置来接收由可移动充电装置 (mobile charger, MC) 传输能量的无线可充电传感器网络 (wireless rechargeable sensor network, WRSN)[3,4]应运而生. 在 WRSN 中, 移动充电器的调度问题, 即如何高效调度 MC 为传感器节点补充能量, 具有重要意义, 是构建具有高能量补充效率、可持续生存 WRSN 的关键.

WRSN 中的充电调度通常包括空间和时间两个维度[5], 其中空间维度指 MC 如何确定一条较优的移动路径, 这涉及如何确定充电顺序以避免节点因等待时间过长而失效, 以及如何尽可能缩短 MC 移动总距离以降低充电成本. 时间维度是指充电时间的分配和调整, 即 MC 需要确定在每个节点停留多长时间为其充电, 从而实现能量的合理分配和整个网络节点之间寿命的平衡. 为了实现有效的充电调度, 已经有许多研究工作被提出, 这些工作大多基于传统的优化方法, 如枚举策略, 近似算法和启发式算法等. 由于 WRSN 充电调度被证明是 NP 难问题[6], 传统方法对于 NP 难问题通常不易得到满足实际需求的优化方案, 因而现有的充电调度往往存在对问题过于简单化或难以很好地适应环境的动态变化等局限.

近年来使用机器学习来优化 WRSN 充电调度问题正迅速成为研究热点, 研究者们已经提出了一些创新性的工作. 例如 Yang 等人[7]充分考虑了全局性的充电调度, 引入了门控循环单元来捕捉时间序列中充电动作之间的关系, 并使用了强化学习 (reinforcement learning, RL) 中的 Actor-Critic 框架来训练该模型. Li 等人[8]利用加权图和 DQN 算法对充电调度问题进行建模, 并提出了相应的充电调度方案. Wei 等人[9]综合考虑了传感器节点能耗变化和节点的地理位置, 提出了一种基于强化学习的无线传感器网络充电策略, 提高了 MC 的自主性. 深度强化学习 (deep reinforcement learning, DRL) 将深度学习与强化学习有机结合, 具有强大的表征能力、决策能力、自主学习能力和独立判断能力[10,11], 对于解决复杂系统的感知和决策问题具有天然的优势, 为解决 WRSN 中移动充电调度的复杂优化问题提供了新的思路. Cao 等人[6]将调度问题转换为 MC 充电过程中的充电奖励最大化问题, 用 MC 的功率和传感器节点的时间窗口对充电奖励进行约束, 并使用 DRL 建模来获得 MC 的充电路径. 然而, 现有的基于机器学习的 WRSN 研究工作只考虑了移动充电调度两个维度中的一个, 即空间维度, 仍然存在性能限制.

本文提出了一种基于 DRL 的新型充电调度方案 SCSD, 从空间和时间两个维度对 WRSN 充电调度中的充电序列和节点充电时长进行优化, 在确保节点能量供应的前提下, 进一步减小了充电成本. 从国内外公开发表的文献来看, 利用 DRL 从空间和时间两个维度优化 WRSN 移动充电调度的研究工作还较少. 本文的主要贡献概括如下.

(1) 基于 DRL 中的 DQN 算法对节点的充电序列进行建模和求解, 生成近似最优的节点充电序列, 再采用适用于具有连续动作空间问题的 DDPG 算法框架, 动态调整充电序列中待充电节点的充电时长. 在提高充电效率的同时, 最大限度避免了后序节点因等待时间过长而缺电失效.

(2) 根据节点剩余生存时间动态调整其发送充电请求的最小能量阈值, 优化 MC 的待充电节点队列长度, 降低调度的计算开销和延迟.

(3) 通过大量实验仿真对 SCSD 的性能进行验证, 仿真实验结果表明: 与现有的几种在线充电策略相比, SCSD 在降低节点失效率的同时, 在充电代价和充电延迟方面具有更好的性能.

本文第 1 节介绍 WRSN 充电调度的相关方法和研究现状. 第 2 节介绍移动能量补充系统模型. 第 3 节介绍本文构建的基于深度强化学习的充电调度模型. 第 4 节通过对比实验, 验证了所提模型的有效性. 最后总结全文.

1 WRSN 充电调度相关工作

基于传统优化方法的 WRSN 移动充电调度研究工作可分为周期性充电调度和按需充电调度两类. 周期性充电通常将为节点充电过程抽象为旅行商 (TSP) 问题, 例如文献 [12−14] 中作者研究了如何生成最短哈密顿回路的问题, 并使 MC 按照生成的最短哈密顿回路为节点进行周期性能量补充. Shi 等人[12]证明了 MC 在每个充电周期中的最优旅行路径是最短哈密顿回路. Han 等人[13]基于 K-means 算法将网络划分为多个簇, MC 通过沿最短哈密

顿循环的反方向移动来访问每个簇中的锚点. Wu 等人[14]将 UAV 能量利用效率最大化问题分解为整数规划和非凸优化问题为其构建充电回路. 在考虑 MC 能量有限的情况下, 魏振春等人[15]提出一种基于精英策略的多目标蚁群优化算法优化 MC 的充电和数据收集策略. Dong 等人[16]考虑节点能耗变化提出了一种以平衡节点间能量消耗为目标的网络模型和一种通信集群与充电集群相结合的周期性调度模型.

为了提高 MC 调度的灵活性, 学者们提出了按需充电模式. He 等人[17]为按需充电模式奠定了理论基础, 节点能量低于阈值时发送充电请求, 使用复杂和高效的最近作业下一步抢占规则 (NJNP) 来确定请求队列中传感器节点的充电顺序. 为了最小化充电延迟, 使待充电节点更快的得到能量补充, Lin 等人[18]提出最小充电延迟问题 (S-MAD) 并将其表述为线性规划问题, 将不断变化的充电方向角划分为一个有限的候选充电方向角集, 并将模型扩展到使用多个 MC 的大规模 WRSN, 有效降低了充电总延迟. Sha 等人[19]根据节点的剩余生存时间将节点分组, 保证每次调度只对剩余能量较低的节点进行充电. 为了提高充电效率, 水九生等人[20]提出了一种按需多节点顺带充电调度方案, 考虑 MC 同时为其覆盖范围内的非请求节点同时充电. Wang 等人[21]使用了一个新的组合充电和收集数据模型 (CRCM) 来建立一个无线传感器网络, MC 用于收集数据和为传感器节点补充能量, 采用 K-means 算法将传感器节点分成不同的簇, MC 从簇中心节点收集数据并采用 NJNP 算法确定充电路径. 针对 TSP 问题在解决能耗不均衡场景的局限性. He 等人[22]基于节点的能耗针对低能量节点构造了一组嵌套的 TSP 问题, 提出了能量同步的概念, 使节点的充电请求序列与旅行商的序列同步, 减少 MC 的行驶距离和节点充电延迟. 考虑到无线传感器网络的主要目的是提供最高的监测质量, Kan 等人[23]在充电的调度中额外考虑了传感器节点的覆盖范围和连通优势, 将监测区域划分为若干个正六边形网格, 根据每个六边形网格的覆盖率和连通性贡献进一步计算充电调度的优先级. 针对三维应用场景, Lin 等人[24]采用无人机 UAV 为节点进行能量补充, 提出一种空间离散化方案以获得三维环境下 UAV 的有限可行充电点集, 以及一种时间离散化方案以确定每个充电点的充电时长. 在考虑多个移动充电器时, 文献 [25,26] 研究了大规模无线传感器网络中多 MC 的运行成本最小化和网络生存时间最大化问题. 考虑到 MC 的个数, Liu 等人[27]提出了叫做 shuttling 的新概念, 在理论上实现了 MC 个数的最小化. Lin 等人[28]在收集充电请求后进行充电序列的优化, 开发了一个节点删除算法删除低效待充电节点, 再执行节点插入算法将低效节点插入避免其失效, 调整充电序列对全局进行优化. Han 等人[29]基于大规模无线传感器网络在充电路径规划中引入异构分簇方案结合采用能量中继传输技术为簇内节点充电.

基于机器学习的 WRSN 充电调度正成为一个研究热点, 一些有效的研究工作已经出现. 例如, Aslam 等人[30]为了提高 WRSN 中的充电效率, 将延长网络寿命和优化节点能耗定义为多目标优化问题, 并基于聚类和强化学习 (SARSA) 的概念, 提出了一种新的数据路由方案 (CSARSA), 智能体通过 SARSA 学习评估的剩余能量状态和动作值集合, 平衡了数据路由过程中的能量消耗, 并提高了网络稳定性. 在使用深度强化学习算法对 MC 充电路径的优化中, 文献 [6] 基于 DQN 算法来获取 MC 的移动路径, 将充电优化过程转换为 DRL 中奖励最大化问题, 求解目标是最大化充电奖励, 为传感器节点设置充电时间窗, 在时间窗和 MC 电池容量的约束下优化传感器节点的充电序列. 文献 [7] 基于强化学习中的 Actor-Critic 框架建立充电调度模型, 可以更快地从候选节点中选择一个最优或接近最优的传感器节点作为下一充电目标. Soni 等人[31]提出基于强化学习的无线传感器网络充电策略, 将状态空间和动作空间离散化, 以 Q-learning 算法为基础将反向链路和移动通信相结合, 通过节点的能量和位置被来评估每次调度中的充电路径, 以延长网络寿命, 提高移动充电的自主性, 但方案中节点的能耗是固定的. 固定的充电路径难以应对节点能耗不均衡的场景.

通过对已有的研究工作的分析可知: 对于 WRSN 移动充电调度问题, 基于传统优化方法的启发式算法通常难以得到满足实际需求的近似最优解. 基于机器学习, 特别是 DRL, 求解该问题是一种更为有效的途径, 但是现有的工作仅从空间维度进行了优化, 性能有待提升. 如何利用 DRL 从空间和时间两个维度实现高效的 WRSN 充电调度优化是本文的研究目标.

2 系统模型与问题描述

本节介绍 WRSN 移动能量补充系统模型, 主要包括 WRSN 网络模型、充电模型、移动充电过程以及传感器节点充电请求阈值的计算. 为便于叙述, 表 1 中列出了本文使用的主要符号.

表 1 主要符号说明

参数名称	描述	参数名称	描述
n	传感器节点数量	r_i	传感器节点i的实时能耗率
T	网络生存时间	R_i	传感器节点i的加权能耗率
T_{cycle}	充电周期时长	T_{th}^{s}	传感器发送充电请求阈值
C_s	传感器电池容量	C_{mc}	MC电池容量
e_r	传感器接收单位数据能耗	V_{mc}	MC移动速度
e_t	传感器发送单位数据能耗	E_{mc}	MC移动单位距离的能耗
$f_{i,j}$	传感器节点i传输到j的数据量	L_{mc}	MC移动距离
p_i	传感器节点i的能耗	ρ	MC对传感器的充电速率

2.1 系统模型

2.1.1 网络模型

整个 WRSN 部署在二维平面区域内如图 1 所示, 由 3 种类型的装置组成: 一个基站 (BS)、一个移动充电设备 (MC) 和大量的传感器节点. 假设 BS 有足够的电量和通信能力, 可以直接向 MC 传输数据. MC 是一种具有自主移动、计算和通信能力的设备, 例如智能小车或移动机器人. MC 有一个容量充足的电池和无线能量传输装置, 它可以通过 BS 快速更换电池[32]. 为实现 WRSN 长期工作, 需要对网络中传感器节点进行有效的能量补充, 使所有节点能够按照预期设定正常工作. 节点在剩余能量低于阈值时向基站发送充电请求, 由基站调度 MC 前往节点位置补充能量.

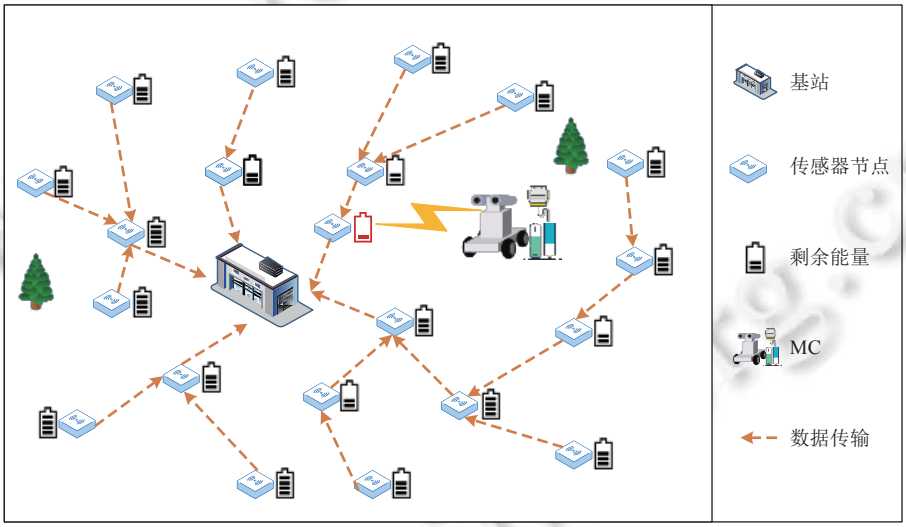


图 1 无线可充电传感器网络模型

(1) 传感器节点设定

网络中部署的 n 个传感器节点固定且位置已知, 根据规划好的传输路径将数据传输至基站. 传感器在接收和转发数据时消耗能量, 并可根据自身数据采集频率和传输速率来预测功耗范围和剩余能量[33], 一旦剩余电量低于预设的阈值就向 BS 发送充电请求. 在实际环境中传感器的能量消耗是动态变化的. 根据文献 [34,35] 对于每个传感器节点 i, 在当前时间步 t 的能耗模型如下:

$$p_i(t)=e_r\sum_{k=1,k\neq i}^{n_r}f_{k,i}(t)+\sum_{j=1,j\neq i}^{n_t}e_t f_{i,j}(t)+e_t f_{i,B}(t) \tag{1}$$

其中, $f_{i,j}(t)$ 和 $f_{i,B}(t)$ 表示当前时间步传感器节点 i 到节点 j 和基站 BS 的数据流, e_r 为节点接收单位数据的能耗, e_t 为节点发送单位数据的能耗, n_r 和 n_t 表示节点 i 接收来自其他 n_r 个节点的数据, 并将数据传输到其他 n_t 个节点.

当前传感器节点 i 的剩余能量计算为:

$$p_r^i(t) = p_r^i(t-1) - p_i(t),\ 1 \leqslant i \leqslant n \tag{2}$$

在当前充电周期内, 节点 i 的剩余能量为 p_r^i, 其每隔时间 Γ 重新计算, 并将 p_r^i 和充电周期标识发送给基站, 节点 i 的实时能耗率为:

$$r_{i,m} = \frac{p_r^i(t-1) - p_r^i(t)}{\Gamma} \tag{3}$$

其中, $r_{i,m}$ 表示接收 m 条剩余能量值信息后基站对节点 i 能耗率的估计, 利用加权平均法将充电周期标识作为计算能耗率的权重[36], 节点 i 的平均能耗率 $R_{i,m}$ 计算为:

$$R_{i,m} = \frac{r_{i,1}t_1 + r_{i,2}t_2 + \ldots + r_{i,m-1}t_{m-1} + r_{i,m}t_m}{t_1 + t_2 + \ldots + t_m} \tag{4}$$

其中, t_m 为传感器节点 i 记录第 m 条剩余能量信息的时间. 由剩余能量 $p_r^i(t)$ 和平均能耗 $R_{i,m}$ 预估节点 i 在当前时间步 t 的剩余生存时间:

$$T_r^i(t) = \frac{p_r^i(t)}{R_{i,m}} \tag{5}$$

当节点 i 的剩余生存时间 $T_r^i(t)$ 低于设定阈值时 i 向 BS 发送充电请求 $< id, l_i, p_c^i(t), T_r^i(t) >$, 其中, $p_c^i(t)$ 为节点当前需要补充的能量:

$$p_c^i(t) = C_s - p_r^i(t) \tag{6}$$

(2) MC 设定

MC 负责维护一个充电服务池, 根据设计的 SCSD 方案选择充电候选节点, 并为传感器进行实时能量补充. 其在 WRSN 区域内以速度 V_{mc} 自由移动, 移动过程中的单位能耗为 E_{mc}, 通过远距离通信直接受基站 BS 调度. 并可通过 GPS 等定位技术实时获取自身位置, MC 只有在到达某个节点位置时为其单独补充能量, 充电功率为 ρ, 其携带电池的最大容量为 C_{mc}. 在为 WRSN 进行充电服务时, MC 的能耗主要分为移动能耗和充电能耗, 在不失一般性的前提下, 我们假设移动能耗与距离呈线性关系, 则在当前时间步 t, MC 剩余能量的计算公式为:

$$C_r(t) = C_r(t-1) - \sum_{i=1}^{n_q} p_i^{\text{charge}}(t-1) - \sum_{i=1}^{n_q} L_{i,i+1}E_{mc} \tag{7}$$

其中, $L_{i,i+1}$ 为 MC 从当前节点移动到下一节点的距离, p_i^{charge} 表示为节点 i 补充的能量, 当 MC 剩余能量小于下一充电回合待充电节点能量需求与自身行驶消耗能量之和时停止为节点补充能量, 从当前位置返回基站充电, MC 自身补充能量的时间忽略不计. MC 在每一个时间步开始时, 首先更新充电服务池中请求节点的剩余能量信息, 然后再进行充电规划. 在剩余能量更新时, 为便于计算假定每个请求节点的能耗速率保持不变. 如此, 在为当前节点进行能量补充时考虑了其余请求节点能量需求的变化.

(3) 基站设定

假设只有一个基站 BS 位于传感器网络中 L_s 处, 作为整个网络的数据存储和处理中心, 并可以持续有效的为 MC 补充能量.

2.1.2 充电模型

(1) 充电请求阈值

传感器节点的充电请求阈值是影响充电性能的重要因素之一, 若阈值设定过大, 节点发送请求时仍剩余较多能量, 一次充电过程为节点补充的能量较少, 导致 MC 能量利用率低下. 若阈值较小, 节点发送请求时剩余能量过低, 则容易来不及充电而饥饿失效. 在节点能耗多变的情况下阈值对充电性能的影响更加明显, 本文根据各个节点的平均能耗率预估剩余生存时间, 提出由节点剩余生存时间确定阈值 T_{th}^s 的方案, 下面对 T_{th}^s 的取值计算展开介绍.

MC 到达之前节点不能饥饿失效, T_{th}^s 应大于 MC 到达待充电节点 i 所需的最短时间:

$$T_{th}^{s} > \frac{L(\mathrm{MC}, n_i)}{V_{mc}} \tag{8}$$

节点间的平均距离为:

$$d = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} d_{i,j}}{n^2} \tag{9}$$

其中, $d_{i,j}$ 为节点 i 到节点 j 之间的欧氏距离, 另外, 我们把节点 i 发送充电请求时需要补充的能量表示为 $p_c^i(t_{th})$, 那么网络中 n_q 个待充电节点的平均等待时间估算为 W_t:

$$\begin{aligned} W_t &= \frac{\left[\frac{d}{V_{mc}} + \frac{2d}{V_{mc}} + \ldots + \frac{n_q d}{V_{mc}} + \frac{p_c^1(t_{th})}{\rho} + \frac{p_c^1(t_{th}) + p_c^2(t_{th})}{\rho} + \ldots + \frac{\sum_{i=1}^{n_q-1} p_c^i(t_{th})}{\rho}\right]}{n_q} \\ &= \frac{(n_q+1)\sum_{i=1}^{n_q}\sum_{j=1}^{n_q} d_{ij}}{2V_{mc}{n_q}^2} + \frac{(n_q-1)p_c^1(t_{th}) + (n_q-2)p_c^2(t_{th}) + \ldots + p_c^{n_q}(t_{th})}{n_q\rho} \end{aligned} \tag{10}$$

其中, $1,2,\ldots,n_q$ 为充电序列中待充电节点的序号, d 为公式 (9) 中的节点间平均距离, ρ 为 MC 对传感器节点的充电速率, 分析可得 T_{th}^{s} 的取值应满足:

$$T_{th}^{s} > \max\left\{\frac{L(\mathrm{MC}, n_i)}{V_{mc}}, W_t\right\} \tag{11}$$

公式 (11) 表示节点在发送充电请求的剩余生存时间阈值应大于节点平均充电等待时间 W_t 和 MC 到该节点的移动时间.

(2) 电池充电模型

根据文献 [37], 目前主流可充电电池接收能量的效率具有边际效应, 即随着充电时间的增加, 电池接收的能量并非线性增加. 在充电过程的前半阶段具有更高的充电效率, 随时间推移, 电池在充电过程中接收能量的速度逐渐降低. 电池的充电模型, 即在充电过程中的能量可由公式 (12) 计算:

$$p_r^i(t') = p_r^i(t) + \int_o^{t_c} \rho(z)\mathrm{d}z \tag{12}$$

其中, t_c 为充电时长, $p_r^i(t')$ 表示经过充电后节点 i 的剩余电量, 多数情况下, 完全充满并不能获得最优的充电效率, $\rho(z)$ 计算如下:

$$\rho(z) = \frac{1}{p} C_s \mathrm{e}^{\frac{-z}{p}} \tag{13}$$

其中, p 为由充电参数决定的时间常数, C_s 为传感器电池最大容量, 将公式 (13) 代入公式 (12) 可得:

$$p_r^i(t') = p_r^i(t) + C_s\left[1 - \mathrm{e}^{-\frac{t_c}{p}}\right] \tag{14}$$

从而可求得在当前时间步 t 对节点进行完全充电所需的时间 t_i^{complete}:

$$t_i^{\mathrm{complete}} = -p\ln\left(1 - \frac{C_s - p_r^i(t)}{C_s}\right) \tag{15}$$

2.2 问题描述

如前所述, 利用 DRL 强大的表征和学习能力来解决 WRSN 充电调度难题是一种极具前景的新途径, 其研究刚刚起步. 从国内外公开发表的文献来看, 现有的基于 DRL 的充电调度方案只考虑了空间维度而没有考虑时间维度, 即没有考虑充电时长的优化调整. 以目前具有代表性的基于 DRL 的充电调度优化方案 RMP-RL 为例. 如图 2

所示, 假设 MC 接收到节点 A, B, E, F, G 的充电请求信息, RMP-RL 采用 DQN 算法根据节点的坐标和 MC 的当前位置规划出一条总行程最短的充电路径 E → G → F → B → A, 依次为上述节点进行充电, 且每次充电都将节点电池充满. 由于 WRSN 中普遍存在着节点能耗的不均匀性, 例如 F 的能耗较快, 可能存在当轮到 MC 为其充电之前该节点能量已经因耗尽的情况, 即节点 F 因缺电而失效. 可见现有的研究工作只考虑了移动充电调度两个维度中的空间维度, 仍然存在性能方面的限制.

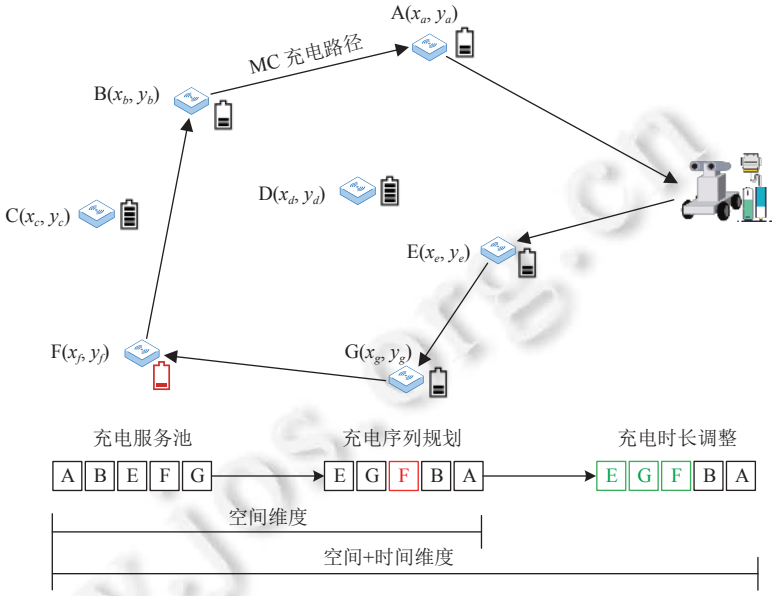


图 2 时空充电调度问题

为解决因充电调度导致的节点失效问题, 本文提出的 SCSD 方法除考虑充电序列调整这一空间维度外, 还重点考虑了充电时长调整这一时间维度, 具体的做法如下: 在图 2 的示例中, 当 MC 基于 DQN 算法规划出一条总行程最短的充电路径 E → G → F → B → A 后, SCSD 将计算所有待充电节点的剩余生存时间, 如果存在剩余生存时间小于等待时间的节点, 例如 F, 则利用适用于具有连续状态空间和行动空间问题的 DDPG 算法, 对充电序列中位于 F 之前的节点 E 和 G 的充电时长进行动态调整, 使得 F 在能量耗尽之前能够得获得充电机会. 此外, 从第 2.1 节公式 (12) 可知, 可充电电池接收能量的效率具有边际效应, 临近充满时边际效益显著降低, 完全充满并不能获得最优的充电效率, 这一理论依据进一步表明了进行充电时长调整以提高整个 WRSN 系统充电效率的必要性和有效性.

3 基于深度强化学习的动态时空充电策略

在本文所提出的 SCSD 充电策略中, 我们假设 MC 在开始阶段位于 BS. 当服务池不为空时, MC 为服务池中的传感器规划充电序列和充电时长并为其充电, 然后返回 BS 为下次调度补充能量. 这样的充电过程称为一次充电任务. 为了最小化节点失效率和充电成本, SCSD 在两个维度上进行协同优化, 即充电顺序和时间. 其中充电序列就是充电调度的空间维度, 为了避免动态能耗率导致传感器失效, MC 需要按上述序列调整每个传感器的充电时长, 进行时间维度的调度优化. 在 SCSD 中, 为构建充电调度优化的 DRL 学习模型, 我们首先给出以下定义.

定义 1 (智能体). 智能体是负责做出和执行充电决策的 MC.

定义 2 (状态). 状态空间集合表示为 $S = S_i\ (i = 1,2,\ldots,n)$, S_i 是包含节点位置、能量需求和平均剩余生存时间的元组, 表示为 $S_i = [l_i, p_c^i(t), T_r^i(t)]$. 其中 l_i 为静态元素, 给出了传感器节点 i 的二维坐标. $p_c^i(t)$, $T_r^i(t)$ 为动态元素, 为当前时间步的能量需求, $T_r^i(t)$ 为节点 i 在当前时间步的剩余生存时间.

定义 3 (动作). 动作表示 MC 如何对观察到的状态做出响应. 在充电序列规划中, 一个动作决定 MC 前往哪个

传感器节点进行能量补充. 在充电时长动态调整中动作是在待充电节点位置的充电时长.

定义 4 (奖励). 奖励是指 MC 在执行策略时对于环境获得的整体回报. 对于充电序列规划, 使用 MC 的总行程长度和失效节点的数量作为奖励信号, MC 的移动距离越短, 一轮充电中失效节点数量越少, MC 获得的奖励值越大. 对于充电时长调整, 使用已充电节点的剩余能量和失效节点数量作为奖励信号, 补充能量后节点的剩余能量越多、失效结点数量越少, 获得奖励值越大.

定义 5 (目标函数). SCSD 方案的总目标为在最小化传感器节点失效率的前提下最大化 MC 充电效率, 该问题是一个多目标优化问题, 其中主要目标是最小化失效传感器节点数量 N_d, 其次是最大化 MC 充电效率, 目标函数设置如下:

$$\min N_d = \sum_{k=1}^{K} N_d^k \tag{16}$$

$$\max \frac{\sum_{k=1}^{K}\sum_{i=1}^{x}\int_0^{t_i}\rho(z)\mathrm{d}z}{E_{mc}\times L_{mc}} \tag{17}$$

约束条件为:

$$\sum_{i\in S} p_c^i(t) + E_{mc}L_{mc} < C_r(t) \tag{18}$$

$$n_{\text{visit}} \in \{0,1\} \tag{19}$$

$$n_s = \begin{cases} 0, & p_r^i(t) = 0 \\ 1, & p_r^i(t) > 0 \end{cases} \tag{20}$$

$$T_{w_i}^{\text{complete}} = \sum_{x=1}^{i-1} ct_x + \sum_{x=1}^{i-1} \frac{L_{x,x+1}}{V_{mc}} \tag{21}$$

$$T_{w_i}^{\text{complete}} < T_r^i(t) \tag{22}$$

目标函数中 K 为充电回合数, N_d 为网络生存周期内总失效节点数量, N_d^k 是每个充电回合中的失效节点数量, $\int_0^{t_i}\rho(z)\mathrm{d}z$ 表示在充电时长 t_i 内节点 i 获得的能量, E_{mc} 为 MC 移动单位距离的能耗, L_{mc} 为 MC 移动距离, 表明最大化 MC 充电效率. 约束条件中公式 (18) 代表待充电节点总能量需求与 MC 移动能耗不能大于 MC 当前剩余能量; 公式 (19) 代表每个传感器节点在一轮充电回合内只能被访问一次; 公式 (20) 指传感器节点具有两种运行状态, 当前剩余能量 $p_r^i(t)=0$ 时被标记为失效节点, 当 $p_r^i(t)>0$ 时正常运行; 公式 (21) 中是 $T_{w_i}^{\text{complete}}$ 是对节点 i 充电序列之前所有传感器进行完全充电时, 节点 i 的等待时间; 公式 (22) 作为是否需要进行充电时长调整的判断条件, 其中 $T_r^i(t)$ 为当前时间步节点 i 的剩余生存时间.

3.1 算法总体步骤

在第 3 节中相关定义的基础上, 采用 DQN 和 DDPG 算法并结合无线传感器网络相关理论, 实现基于深度强化学习的时空充电调度算法.

算法总体步骤如下.

步骤 1. 将 WRSN 运行时间划分为多个连续的充电周期, 在每个周期开始时传感器节点通过平均能耗率 R_i 计算自身剩余生存时间 $T_r^i(t)$, 当 $T_r^i(t) < T_{th}^s$ 时发送充电请求 $< id, l_i, p_c^i(t), T_r^i(t) >$, 其中 id 为传感器节点的唯一标识, l_i 为其位置坐标, $p_c^i(t)$, $T_r^i(t)$ 分别为传感器发送充电请求时的能量需求和剩余生存时间. 充电请求发出后被存入充电服务池 S.

步骤 2. 当充电服务池 S 不为空时, MC 为 S 中的每个节点计算充电调度方案. 首先根据 DQN 算法为充电调度中的充电序列规划问题建模求解, 将充电节点选择转换为离散的动作空间中求解 DRL 奖励最大化问题, 根据节点的位置, 剩余能量, 和失效节点数量设计奖励函数, 使 MC 优先为距离近, 剩余能量低的传感器补充能量, 求出充

电序列 CS.

步骤 3. MC 根据求出的充电序列 CS 依次访问待充电节点, 首先遍历 CS 各个节点计算完全充电所需时间 ct_i, 对于节点 i, 比较前 $i-1$ 个节点完全充电所需时长, 若其大于等于自身剩余生存时间, 则采用 DDPG 算法对前 $i-1$ 的充电时长进行调整, 若小于自身剩余生存时间则为前 $i-1$ 个节点进行完全充电.

步骤 4. 完成一次充电回合后, MC 返回 BS 补充能量准备下一回合调度, 并在新的充电回合重复步骤 1 到 3.

上述步骤中, 步骤 2 对充电序列的优化和步骤 3 对充电时长的优化是整个 SCSD 算法的核心, 下面分别对其进行详细介绍.

3.2 基于 DQN 的充电序列优化

作为充电调度的空间维度, 充电序列决定了 MC 的充电性能和节点是否能及时得到能量补充. 为了优化充电序列, 在充电调度中应满足以下约束条件.

1) 待充电节点的总能量需求不大于 MC 当前剩余能量, 即满足约束公式 (18).

2) 在一轮充电调度中, 同一节点不可被访问两次, 以避免 MC 的低效往返移动, 即满足约束公式 (19).

3) 每个传感器节点的状态分为正常工作和失效, 当节点能量耗尽时被标记为失效节点, 即满足约束公式 (20).

由于充电序列规划问题具有离散的动作空间, 因此我们采用一种基于值的 DRL 算法, 即 DQN 算法[6,38]将充电序列优化转换为 DQN 中的奖励最大化问题, 如图 3 所示, 其中 MC 的动作空间定义为 $A=\{a_1,a_2,\ldots,a_{n+1}\}$. 当 $A=a_{n+1}$ 时 MC 返回 BS 为自身补充能量; $A=a_i$ $(i=1,2,\ldots,n)$ 表示 MC 为节点 i 补充能量. 如前所述, WRSN 与 MC 分别对应 DRL 模型中的环境和智能体, 并相互作用. 在充电调度过程中每执行一次动作环境都会反馈给 MC 一个奖励, 并更新到新的环境状态. 当 MC 到达最终状态时计算总奖励, 并通过训练 Q 函数使 MC 尽可能获得更多的奖励, 从而构造 MC 的最优充电序列.

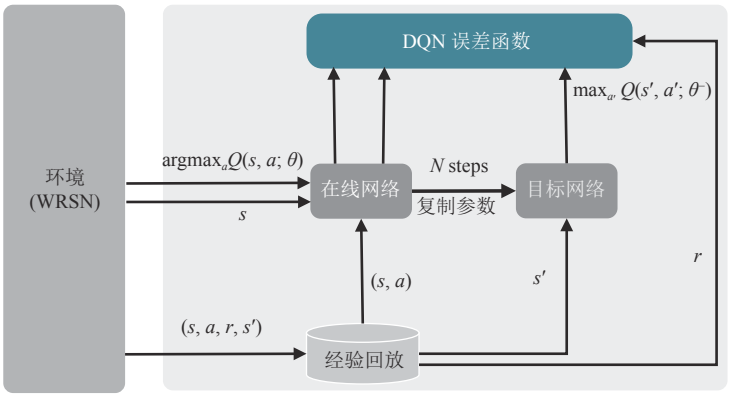


图 3 基于 DQN 的充电序列优化

在充电序列规划中, 奖励函数的设计需要考虑 3 个因素: 传感器能量因素 E_i, 传感器位置因素 D_i 和传感器失效因素 N_i. 3 个重要因素如下所示:

$$\begin{cases} E_i=\sum_{i=1}^{x}\dfrac{R_{i,m}}{p_r^i} \\ D_i=-\sum_{i=1}^{x}L_{i,i+1} \\ N_i=-\alpha N_d^i \end{cases} \tag{23}$$

其中, $R_{i,m}$ 为节点实时能耗率, p_r^i 为节点 i 的剩余能量, 表明在能量因素 E_i 中鼓励算法优先选择能耗高且剩余能量较低的节点补充能量. $L_{i,i+1}$ 为当前待充电节点与下一待充电节点的距离, 表明距离因素优先选择距离近的节点, 降低 MC 移动距离进而提高充电效率. α 为惩罚系数, 当在状态 s_t 执行动作 a_t 后造成充电序列中 N_d^i 个节点失效时

反馈给算法一个负奖励.

结合上述影响充电质量的 3 个因素, 奖励函数设计如下, 其中 K 为传感器网络生存周期内充电调度回合数, 充电序列规划目标为总奖励最大化.

$$R_s=\sum_{k=1}^{K}\left(\sum_{i=1}^{x}\frac{R_{i,m}}{p_r^i}-\sum_{i=1}^{x}L_{i,i+1}-\alpha N_d^i\right) \tag{24}$$

在训练过程中定义每个 episode 为 MC 的一个充电回合, 其中的每个 step 为 MC 执行的充电动作. DQN 的目标是最大化未来奖励, 每个时间步的折扣因子为 γ, 则在时间步 t 的未来奖励 $R=\sum_{t'}^{T}\gamma^{t'-t}r_{t'}$, 定义动作价值函数 $Q^*=(s,a)$ 为所有动作价值中的最大值. 对于下一状态 s' 所有可能存在的动作 a', 当前的最优策略是选择一个使目标函数最大的动作. 首先将当前状态 s_t 发送到 Q 网络并计算每个动作的 Q 值, 采用 ε 贪婪策略选择 Q 值最大的动作, MC 执行动作 a_t 后计算奖励 r_t, 再更新当前状态为 s_{t+1}.

$$Q^*(s,a)=\mathbb{E}_{s'\sim\varepsilon}\left[r+\gamma\max_{a'}Q^*(s',a')|s,a\right] \tag{25}$$

Q 网络为带有权重 θ 的神经网络函数逼近器 $Q(s,a;\theta)\approx Q^*(s,a)$ 通过最小化损失函数 $L_i(\theta_i)$ 来训练 Q 网络, 损失函数的梯度计算公式如下:

$$\nabla_{\theta_i}L_i(\theta_i)=\mathbb{E}_{s,a\sim\rho(\cdot);s'\sim\varepsilon}\left[(r+\gamma\max_{a_{t+1}}Q(s',a';\theta_{i-1})-Q(s,a;\theta_i))\nabla_{\theta_i}Q(s,a;\theta_i)\right] \tag{26}$$

算法 1. 基于 DQN 的充电序列优化.

输入: 充电服务池 S;
输出: 充电路径 CS.

1. 初始化经验复用池 $\mathcal{D}$;
2. 初始化动作对应的 Q 值为随机值;
3. **for** episode = 1:M **do**
4. 初始化部分充电路径 CS_1;
5. **for** t=1:T **do**
6. 以 ε 贪婪策略选择动作 $a_t=\max_a Q^*(s_t,a;\theta)$;
7. 在模拟器中执行动作 a_t 并观察对应奖励 r_t;
8. 添加 a_t 到部分充电路径 $CS_{t+1}\leftarrow(CS_t,a_{t+1})$;
9. 在复用池 $\mathcal{D}$ 中存储样本 (s_t,a_t,r_t,s_{t+1});
10. 随机抽取小批量 (mini-batch) 的样本 (s_j,a_j,r_j,s_{j+1});
11. 令 $y_i=\begin{cases}r_j, \text{ if } s_{i+1}\\ r_j+\gamma\max\limits_{a'}Q(s_{i+1},a';\theta), \text{ else}\end{cases}$
12. 根据公式 (26) 在 $(y_i-Q(s_{i+1},a';\theta))^2$ 执行梯度下降;
13. **end for**
14. **end for**

3.3 基于 DDPG 的充电时长优化

在完成充电序列规划后得到一条覆盖所有请求节点的路径. 此时 MC 沿路径访问所有待充电节点的能量损失最小. 但若充电序列中的所有节点都被完全充电, 可能会造成序列中剩余能量较低的部分节点由于没有及时充电而饥饿失效. 因此为了平衡节点间的能量需求, 需要动态调整它们的充电时间. 与充电序列规划不同, 充电时间调整中的动作是向一个节点补充多少能量, 为连续的动作空间, 因此需要单独对其建模求解. 我们采用 DDPG (deep deterministic policy gradient) 算法框架[39]针对这一针对连续动作空间的策略学习设计解决方案, 基于 DDPG 的充电时长调整网络结构如图 4 所示.

DDPG 结合了 Actor-Critic 和 DQN, 在 Actor-Critic 结构的基础上采用 DQN 的双网络结构和经验回放机制, 解决 Actor-Critic 难收敛和 DQN 不能解决连续动作空间的问题. Actor 网络直接输出确定性的动作, 即节点的充电时长, 再由 Critic 网络对这一动作进行评判, 进而更新 Actor 网络. 其包含 4 个网络结构: Actor 网络负责参数 θ 的迭代更新, 并根据当前状态 S 选择动作 A, 用于和环境交互生成 S', R; 目标 Actor 网络负责根据经验回放池中采样的下一状态 S' 选择最优下一动作 A', 其中网络参数 θ' 定期从 θ 复制; 目标 Critic 网络负责价值网络参数 w 的迭代更新并计算当前 Q 值 $Q_w(s,a)$; 目标 Q 值 $Q_{\text{target}}=r+\gamma Q_{\bar{w}}(s',a')$, 网络参数 $\bar{w}$ 定期从 w 复制, 使用神经网络来模拟策略函数和 Q 函数, 在训练中得到最大化系统的预期奖励.

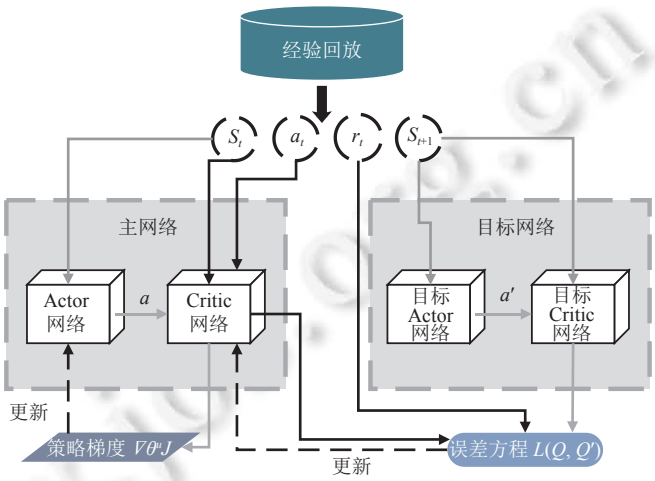


图 4 基于 DDPG 的充电时长调整

在 DDPG 中神经网络的训练集为 (s_t,a_t,r_t,s_{t+1}), 对于充电时长调整, 状态空间包括节点的充电序列、剩余生存时间和失效节点数量, 动作空间为待充电序列中节点的充电时长, 充电时长调整分为以下步骤.

1) 首先根据公式 (21) 计算对节点 i 充电序列之前节点进行完全充电时的节点 i 的充电等待时间 $T_{w_i}^{\text{complete}}$.

2) 若满足约束公式 (22), 则对前 i–1 个节点进行完全充电, 此时待充电节点的总能量需求要小于 MC 当前剩余能量.

3) 若不满足约束公式 (22), 则对前 i–1 个节点的充电时长进行动态调整, 使得节点 i 尽可能地在能量耗尽之前得到补充.

4) 在最小化节点失效率的前提下, 尽可能多的为节点补充能量.

在 DDPG 的更新阶段, 首先从经验复用池 $\mathcal{D}$ 中抽取若干经验通过目标网络得到目标奖励值 y_i, 具体为将下一状态向量 s_{t+1} 放入目标策略网络得到动作 a_{t+1} 之后将目标动作和下一状态向量共同作为目标价值网络的输入得到目标值 Q' 后根据公式:

$$y_i=r_i+\gamma Q'(s_{i+1},\mu'(s_{i+1}|\theta^{Q'})) \tag{27}$$

得出目标奖励值 y_i 后更新 Critic 网络, 首先输入 s_i,a_i 得到实际 Q 值, 根据误差方程:

$$L=\frac{1}{N}\sum_i(y_i-Q(s_i,a_i|\theta^Q)^2) \tag{28}$$

DDPG 中的策略网络为确定性更新, 得出的为确定的充电时长, 根据策略梯度 $\nabla_{\theta^\mu}J$ 确定 Actor 网络的更新方向:

$$\nabla_{\theta^\mu}J=\frac{1}{N}\sum_i\nabla_a Q(s,a|\theta^Q)|_{s=s_i,a=\mu(s_i)}\nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s_i} \tag{29}$$

根据以上描述, 基于 DDPG 的充电时长调整算法如下.

算法 2. 基于 DDPG 的充电时长调整.

输入: 充电路径 CS;

输出: 各节点充电时长 t_c .

1. 随机初始化策略网络参数 θ^{μ} 和价值网络参数 θ^{Q} ;
2. 随机初始化目标策略网络参数 $\theta^{\mu'}$ 和目标价值网络参数 $\theta^{Q'}$;
3. 初始化经验复用池 $\mathcal{D}$;
4. **for** episode = 1:M **do**
5. 初始化动作探索的随机噪声 $\mathcal{N}$;
6. 得到初始化状态 s_1 ;
7. **for** t=1:T **do**
8. 根据在线策略网络和探索随机噪声选择动作 $a_t=\mu(s_t|\theta^{\mu})+\mathcal{N}_t$;
9. 执行动作 a_t 从环境得到奖励 r_t 和下一状态 s_{t+1} ;
10. 将 (s_t,a_t,r_t,s_{t+1}) 存储到经验复用池 $\mathcal{D}$;
11. 从 $\mathcal{D}$ 中随机采样 N 个 (s_i,a_i,r_i,s_{i+1}) ;
12. 通过公式 (27) 得到目标值 y_i ;
13. 通过公式 (28) 最小化损失函数更新 Q 网络;
14. 通过公式 (29) 更新策略网络;
15. 更新目标网络:

$$\theta^{Q'} \leftarrow \tau\theta^{Q}+(1-\tau)\theta^{Q'}$$

$$\theta^{\mu'} \leftarrow \tau\theta^{\mu}+(1-\tau)\theta^{\mu'}$$

16. **end for**
17. **end for**

4 实验分析

本节将进行仿真实验来评估所提充电策略的有效性. 首先基于 Python 建立深度强化学习框架对 SCSD 进行性能分析并与 RMP-RL[6]、NJNP[17]、FCFS[40]进行性能对比. 仿真实验在 200 m×200 m 的正方形区域内随机分布了 25–200 个传感器节点, 其余默认参数设置如表 2 所示.

表 2 实验参数

参数	数值
网络区域	200 m×200 m
网络运行时间	72 000 s
C_{mc} , MC电池容量	1 000 J
C_s , 传感器电池容量	10 J
V_{mc} , MC移动速度	5 m/s
ρ , MC充电速率	200 mJ/s
E_{mc} , MC移动单位距离能耗	4 mJ
e_r , 传感器接收单位数据能耗	0.5 mJ
e_t , 传感器发送单位数据能耗	0.4 mJ

4.1 评价指标

我们通过 3 个关键指标对比分析 SCSD 对 WRSN 的充电调度优化效果.

1) 节点失效率: 当节点剩余能量为 0 时被标记为失效, 节点失效率被定义为失效节点与请求节点总数之比, 是评价充电策略的重要指标之一, 节点失效率越小则充电性能越高, 网络生存周期越长.

2) 充电延迟: 定义为节点从发送充电请求到 MC 开始为节点补充能量之时的时间间隔.

3) 充电代价: 即 MC 行驶总距离, 定义为网络生存周期内所有充电调度回合内 MC 移动距离之和, 充电路径长度越短则 MC 的能量利用率越高, 充电性能和效率越高.

4.2 实验参数

根据对实际应用场景的考虑和现有文献的参考, 实验中设置的参数如表 2 所示.

4.3 节点能耗速率对性能的影响

节点能耗速率越高意味着节点需要越快的补充能量, 节点越容易饥饿失效. 充电策略能否应对传感器能耗多变的环境, 是判断一种充电策略是否符合实际环境的一项重要指标. 为分析其对充电策略性能的影响, 本组实验把传感器节点数据产生的平均时间间隔调整为 30 s、40 s、50 s、60 s、70 s、80 s. 由于数据产生时间的变化, 节点的能耗速率会随之变化. 得到的实验结果如图 5 所示.

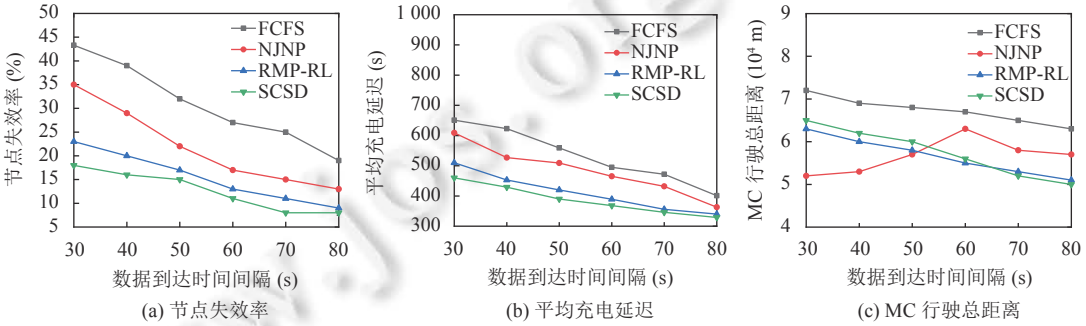


图 5 节点能耗速率对充电性能的影响

从图 5(a) 和图 5(b) 可以看出, 随着数据产生的平均到达时间间隔增大, 4 种策略中节点失效率和充电延迟都呈下降趋势, 原因是节点能耗率下降, 待充电节点减少. 而从图 5(c) 中可以看到, 随着数据产生的平均到达时间间隔增大 NJNP 策略中 MC 总移动距离先增大后减小, 这是由于当节点能耗速率过高时请求节点数量多, 从而导致 MC 来不及相应距离较远的节点, 此时优先为 MC 周围节点补充能量, 以至于移动距离较短但节点死亡率较高, 当节点能耗率下降时, 请求节点数量减少, MC 可以为距离较远的节点补充能量, 故移动距离增加. 随着节点能耗率降低充电需求减少因此 MC 行驶距离减少. 在 SCSD 策略中随着节点能耗速率的减小, MC 的总移动距离减少, 原因是无论请求节点数量如何, SCSD 从全局考虑优化充电序列和充电时长, 为整个网络中的节点补充能量, 请求节点数量多, 则 MC 移动距离大. 在数据到达时间间隔为 50 s 之前 SCSD 的 MC 总移动距离大于 NJNP 和 RMP-RL, 这是由于 SCSD 策略为最小化节点死亡数量, 为节点进行部分充电, 所以单个节点再次请求补充能量的概率增加, 使 MC 的移动距离有所增加但从图 5(a) 中可以看出, SCSD 策略的节点失效率始终低于其他 3 种策略, 保证了网络的正常运行. 在数据到达时间间隔大于 70 s 时在 MC 移动距离上 SCSD 也取得了比其他 3 种策略更好的结果.

4.4 节点数量对性能的影响

在本组实验中我们将节点数量由 25 逐渐增加到 200, 节点数量越多则在同一时段内需要补充能量的节点越多, 由此测试各充电策略的性能, 得到的实验结果如图 6 所示.

分析图 6(a) 可得, 当网络中节点数量较少时, MC 能够及时响应待充电节点并为其补充能量, 此时 4 种策略节点失效率相差不大. 随着待充电节点数量增加, MC 充电负荷增加, 部分节点得不到及时的能量补充, 故节点失效率增加, 但 SCSD 策略的节点失效率始终低于另外 3 种策略. 如图 6(b) 所示, 由于 FCFS 和 NJNP 的充电序列并非

最优, 故节点数量多时平均充电延迟较高, 而 SCSD 通过调整充电时长相比 RMP-RL 进一步减少了充电延迟. 通过图 6(c) 可得, NJNP 策略中 MC 的平均移动距离先增大后减小. 这是由于当请求节点数量达到一定程度时, NJNP 优先为距离 MC 较近的节点充电, 不会响应距离较远的待充电节点, 使得 MC 移动距离减小. SCSD 在请求节点数量过多时为节点进行部分充电, 使得节点再次请求充电的概率增大, 故比 RMP-RL 移动距离略高, 但 SCSD 策略尽可能保证了节点的正常工作.

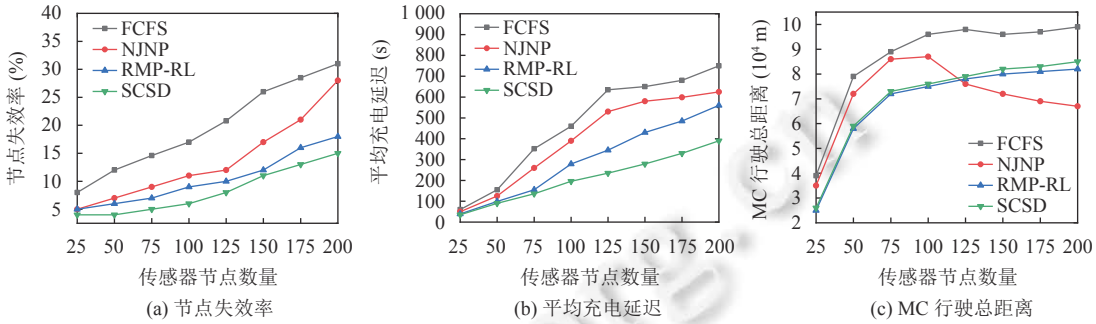


图 6 节点数量对充电性能的影响

4.5 MC 充电速率对性能的影响

本组实验通过改变 MC 对传感器节点的充电速率观察分析对整个传感网络充电调度的影响, 充电速率 ρ 从 100 mJ/s 增加到 350 mJ/s, 得到结果如图 7 所示.

分析图 7(a) 和图 7(b) 可得, 由于充电速率增加使得各个节点的充电时长缩短, MC 在单位时间内可以服务更多的待充电节点从而更快响应下一待充电节点, 因此 4 种策略中节点失效率和平均充电延迟都呈减小趋势, 且 SCSD 在结果上始终优于其他 3 种策略. 图 7(c) 显示了随着充电速率的增加, 4 种策略中 MC 总移动距离都呈增加趋势, 这是由于充电效率越大 MC 可响应的节点越多, MC 来回往返移动越多, 移动总距离就越大. SCSD 和 RMP-RL 两种策略在 MC 行驶总距离上相差不多, 两种策略在充电速率较低时 MC 移动距离高于 NJNP 策略, 这是由于充电速率较低时 NJNP 中 MC 仅能为部分节点服务, 故总移动距离小, SCSD 和 RMP-RL, 面向所有节点进行充电服务, 故总移动距离高. 且 SCSD 在有需要时为节点进行部分充电, 在降低节点失效率的前提下 MC 总移动距离略高于 RMP-RL, 而在充电速率较高时 SCSD 策略的 MC 移动距离低于其他策略, 能量利用率更高, 展现了采用强化学习同时对时间和空间进行全局充电调度优化的有效性.

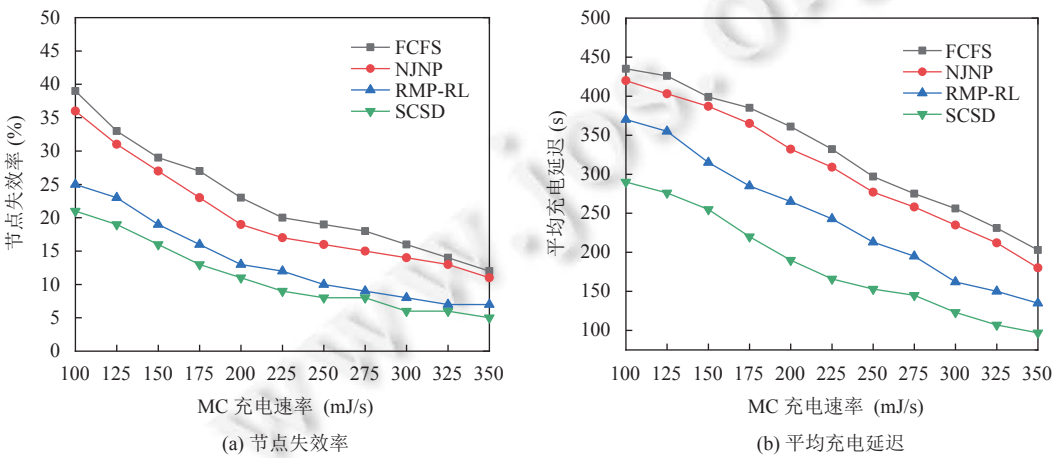
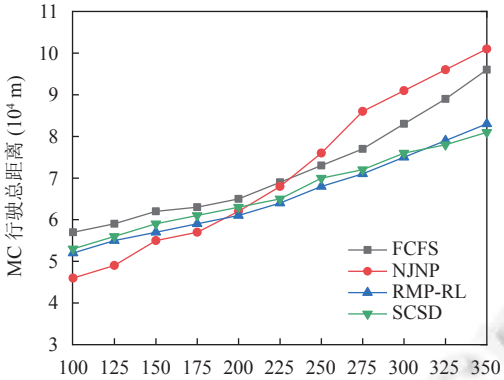


图 7 MC 充电速率对充电性能的影响



(c) MC 行驶总距离

图 7 MC 充电速率对充电性能的影响 (续)

5 总 结

本文研究了 WRSN 的移动充电调度优化问题, 提出了基于 DRL 的动态时空充电方法 SCSD. 首先, SCSD 方法以减少充电代价、提高充电效率为目标, 采用 DQN 算法得到近似最优的充电序列. 然后, 利用 DDPG 算法框架基于节点的剩余生存时间来动态调整序列中待充电节点的充电时长, 从而很好地适应了实际环境中传感器节点能量消耗的动态性, 避免节点因过长的充电等待而缺电失效. 仿真实验中, 我们以节点失效比率、充电延迟和充电代价为主要评价指标, 大量的仿真实验结果表明: 与现有基于深度强化学习的 RMP-RL 算法以及基于传统优化方法的 NJNP 和 FCFS 充电调度方案相比, SCSD 均表示出更好的性能: 能够显著降低节点失效比率, 有效降低充电成本并提高充电响应的及时性.

References:

[1] Kurs A, Karalis A, Moffatt R, Joannopoulos JD, Fisher P, Soljačić M. Wireless power transfer via strongly coupled magnetic resonances. Science, 2007, 317(5834): 83–86. [doi: 10.1126/science.1143254]

[2] Zhang Z, Pang HL, Georgiadis A, Cecati C. Wireless power transfer—An overview. IEEE Trans. on Industrial Electronics, 2019, 66(2): 1044–1058. [doi: 10.1109/TIE.2018.2835378]

[3] Hu C, Wang Y, Wang H. Survey on charging programming in wireless rechargeable sensor networks. Ruan Jian Xue Bao/Journal of Software, 2016, 27(1): 72–95 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4883.htm [doi: 10.13328/j.cnki.jos.004883]

[4] He SB, Chen JM, Jiang FC, Yau DKY, Xing GL, Sun YX. Energy provisioning in wireless rechargeable sensor networks. IEEE Trans. on Mobile Computing, 2013, 12(10): 1931–1942. [doi: 10.1109/TMC.2012.161]

[5] Wang C, Li J, Ye F, Yang YY. Improve charging capability for wireless rechargeable sensor networks using resonant repeaters. In: Proc. of the 35th IEEE Int'l Conf. on Distributed Computing Systems. Columbus: IEEE, 2015. 133–142. [doi: 10.1109/ICDCS.2015.22]

[6] Cao XB, Xu WZ, Liu XX, Peng J, Liu T. A deep reinforcement learning-based on-demand charging algorithm for wireless rechargeable sensor networks. Ad Hoc Networks, 2021, 110: 102278. [doi: 10.1016/j.adhoc.2020.102278]

[7] Yang MY, Liu NB, Zuo L, Feng Y, Liu MH, Gong HG, Liu M. Dynamic charging scheme problem with actor-critic reinforcement learning. IEEE Internet of Things Journal, 2021, 8(1): 370–380. [doi: 10.1109/JIOT.2020.3005598]

[8] Li X, Jin M. Charger scheduling optimization framework. In: Proc. of the 18th IEEE Int'l Symp. on Network Computing and Applications. Cambridge: IEEE, 2019. 1–8. [doi: 10.1109/NCA.2019.8935036]

[9] Wei ZC, Liu F, Lyu Z, Ding X, Shi L, Xia CK. Reinforcement learning for a novel mobile charging strategy in wireless rechargeable sensor networks. In: Proc. of the 13th Int'l Conf. on Wireless Algorithms, Systems, and Applications. Tianjin: Springer, 2018. 485–496. [doi: 10.1007/978-3-319-94268-1_40]

[10] Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, Silver D, Wierstra D. Continuous control with deep reinforcement learning. In:

Proc. of the 4th Int'l Conf. on Learning Representations. San Juan: ICLR, 2016.

[11] Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, Riedmiller M. Playing atari with deep reinforcement learning. arXiv:1312.5602, 2013.

[12] Shi Y, Xie LG, Hou YT, Sherali HD. On renewable sensor networks with wireless energy transfer. In: Proc. of the 2011 IEEE INFOCOM. Shanghai: IEEE, 2011. 1350–1358. [doi: 10.1109/INFCOM.2011.5934919]

[13] Han GJ, Yang X, Liu L, Zhang WB. A joint energy replenishment and data collection algorithm in wireless rechargeable sensor networks. IEEE Internet of Things Journal, 2018, 5(4): 2596–2604. [doi: 10.1109/JIOT.2017.2784478]

[14] Wu PF, Xiao F, Sha C, Huang HP, Sun LJ. Trajectory optimization for UAVs' efficient charging in wireless rechargeable sensor networks. IEEE Trans. on Vehicular Technology, 2020, 69(4): 4207–4220. [doi: 10.1109/TVT.2020.2969220]

[15] Wei ZC, Sun RH, Lyu ZW, Han JH, Shi L, Xu JY. Path planning algorithm for WCE with joint energy replenishment and data collection based on multi-objective optimization. Journal on Communications, 2018, 39(10): 22–33 (in Chinese with English abstract). [doi: 10.11959/j.issn.1000-436x.2018216]

[16] Dong Y, Li SY, Bao GJ, Wang CY. An efficient combined charging strategy for large-scale wireless rechargeable sensor networks. IEEE Sensors Journal, 2020, 20(17): 10306–10315. [doi: 10.1109/JSEN.2020.2990641]

[17] He L, Kong LH, Gu Y, Pan JP, Zhu T. Evaluating the on-demand mobile charging in wireless sensor networks. IEEE Trans. on Mobile Computing, 2015, 14(9): 1861–1875. [doi: 10.1109/TMC.2014.2368557]

[18] Lin C, Yang ZW, Dai HP, Cui LX, Wang L, Wu GW. Minimizing charging delay for directional charging. IEEE/ACM Trans. on Networking, 2021, 29(6): 2478–2493. [doi: 10.1109/TNET.2021.3095280]

[19] Sha C, Sun Y, Malekian R. Research on cost-balanced mobile energy replenishment strategy for wireless rechargeable sensor networks. IEEE Trans. on Vehicular Technology, 2020, 69(3): 3135–3150. [doi: 10.1109/TVT.2019.2962877]

[20] Shui JS, Xu XH. An on-demand passer-by charging scheme based on multi-node charging model. Acta Electronica Sinica, 2021, 49(2): 346–353 (in Chinese with English abstract). [doi: 10.12263/DZXB.20200363]

[21] Wang YH, Dong Y, Li SY, Huang RY, Shang YH. A new on-demand recharging strategy based on cycle-limitation in a WRSN. Symmetry, 2019, 11(8): 1028. [doi: 10.3390/sym11081028]

[22] He L, Fu LK, Zheng LK, Gu Y, Cheng P, Chen JM, Pan JP. ESync: An energy synchronized charging protocol for rechargeable wireless sensor networks. In: Proc. of the 15th ACM Int'l Symp. on Mobile ad Hoc Networking and Computing. Philadelphia: ACM, 2014. 247–256. [doi: 10.1145/2632951.2632970]

[23] Kan YP, Chang CY, Kuo CH, Roy DS. Coverage and connectivity aware energy charging mechanism using mobile charger for WRSNs. IEEE Systems Journal, 2022, 16(3): 3993–4004. [doi: 10.1109/JSYST.2021.3109056]

[24] Lin C, Guo CY, Dai HP, Wang L, Wu GW. Near optimal charging scheduling for 3-D wireless rechargeable sensor networks with energy constraints. In: Proc. of the 39th IEEE Int'l Conf. on Distributed Computing Systems. Dallas: IEEE, 2019. 624–633. [doi: 10.1109/ICDCS.2019.00068]

[25] Xu WZ, Liang WF, Lin XL, Mao GQ. Efficient scheduling of multiple mobile chargers for wireless sensor networks. IEEE Trans. on Vehicular Technology, 2016, 65(9): 7670–7683. [doi: 10.1109/TVT.2015.2496971]

[26] Wang K, Wang L, Lin C, Obaidat MS, Alam M. Prolonging lifetime for wireless rechargeable sensor networks through sleeping and charging scheduling. Int'l Journal of Communication Systems, 2020, 33(8): e4355. [doi: 10.1002/dac.4355]

[27] Liu T, Wu BJ, Wu HY, Peng J. Low-cost collaborative mobile charging for large-scale wireless sensor networks. IEEE Trans. on Mobile Computing, 2017, 16(8): 2213–2227. [doi: 10.1109/TMC.2016.2616309]

[28] Lin C, Zhou JZ, Guo CY, Song HB, Wu GW, Obaidat MS. TSCA: A temporal-spatial real-time charging scheduling algorithm for on-demand architecture in wireless rechargeable sensor networks. IEEE Trans. on Mobile Computing, 2018, 17(1): 211–224. [doi: 10.1109/TMC.2017.2703094]

[29] Han GJ, Guan HF, Wu JW, Chan S, Shu L, Zhang WB. An uneven cluster-based mobile charging algorithm for wireless rechargeable sensor networks. IEEE Systems Journal, 2019, 13(4): 3747–3758. [doi: 10.1109/JSYST.2018.2879084]

[30] Aslam N, Xia KW, Hadi MU. Optimal wireless charging inclusive of intellectual routing based on SARSA learning in renewable wireless sensor networks. IEEE Sensors Journal, 2019, 19(18): 8340–8351. [doi: 10.1109/JSEN.2019.2918865]

[31] Soni S, Shrivastava M. Novel wireless charging algorithms to charge mobile wireless sensor network by using reinforcement learning. SN Applied Sciences, 2019, 1(9): 1052. [doi: 10.1007/s42452-019-1091-2]

[32] Hu C, Wang Y. Schedulability decision of charging missions in wireless rechargeable sensor networks. In: Proc. of the 11th Annual IEEE Int'l Conf. on Sensing, Communication, and Networking. Singapore: IEEE, 2014. 450–458. [doi: 10.1109/SAHCN.2014.6990383]

[33] D'Arienzo M, Iacono M, Marrone S, Nardone R. Estimation of the energy consumption of mobile sensors in WSN environmental

monitoring applications. In: Proc. of the 27th Int'l Conf. on Advanced Information Networking and Applications Workshops. Barcelona: IEEE, 2013. 1588–1593. [doi: 10.1109/WAINA.2013.33]

[34] Xie LG, Shi Y, Hou YT, Sherali HD. Making sensor networks immortal: An energy-renewal approach with wireless power transfer. IEEE/ACM Trans. on Networking, 2012, 20(6): 1748–1761. [doi: 10.1109/TNET.2012.2185831]

[35] Hou YT, Shi Y, Sherali HD. Rate allocation and network lifetime problems for wireless sensor networks. IEEE/ACM Trans. on Networking, 2008, 16(2): 321–334. [doi: 10.1109/TNET.2007.900407]

[36] Zhu JQ, Feng Y, Sun HZ, Liu M, Zhang ZN. Energy starvation avoidance mobile charging for wireless rechargeable sensor networks. Ruan Jian Xue Bao/Journal of Software, 2018, 29(12): 3868–3885 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5315.htm [doi: 10.13328/j.cnki.jos.005315]

[37] Zhao CX, Zhang HJ, Chen FL, Chen SG, Wu CZ, Wang TC. Spatiotemporal charging scheduling in wireless rechargeable sensor networks. Computer Communications, 2020, 152: 155–170. [doi: 10.1016/j.comcom.2020.01.037]

[38] Chu M, Li H, Liao XW, Cui SG. Reinforcement learning-based multiaccess control and battery prediction with energy harvesting in IoT systems. IEEE Internet of Things Journal, 2019, 6(2): 2009–2020. [doi: 10.1109/JIOT.2018.2872440]

[39] Qiu CR, Hu Y, Chen Y, Zeng B. Deep deterministic policy gradient (DDPG)-based energy harvesting wireless communications. IEEE Internet of Things Journal, 2019, 6(5): 8577–8588. [doi: 10.1109/JIOT.2019.2921159]

[40] Panwalkar SS, Iskander W. A survey of scheduling rules. Operations Research, 1977, 25(1): 45–61. [doi: 10.1287/opre.25.1.45]

附中文参考文献:

[3] 胡诚, 汪芸, 王辉. 无线可充电传感器网络中充电规划研究进展. 软件学报, 2016, 27(1): 72–95. http://www.jos.org.cn/1000-9825/4883.htm [doi: 10.13328/j.cnki.jos.004883]

[15] 魏振春, 孙仁浩, 吕增威, 韩江洪, 石雷, 徐俊逸. 联合充电和数据收集的WCE多目标路径规划算法. 通信学报, 2018, 39(10): 22–33. [doi: 10.11959/j.issn.1000-436x.2018216]

[20] 水九生, 徐向华. 一种基于多节点充电模型的按需顺带充电方案. 电子学报, 2021, 49(2): 346–353. [doi: 10.12263/DZXB.20200363]

[36] 朱金奇, 冯勇, 孙华志, 刘明, 张兆年. 无线可充电传感器网络中能量饥饿避免的移动充电. 软件学报, 2018, 29(12): 3868–3885. http://www.jos.org.cn/1000-9825/5315.htm [doi: 10.13328/j.cnki.jos.005315]

王艺均(1997－), 男, 博士生, 主要研究领域为无线传感器网络, 强化学习.

刘明(1972－), 男, 博士, 教授, 博士生导师, 主要研究领域为无线传感器网络, 智能感知, 深度学习.

冯勇(1975－), 男, 博士, 教授, 博士生导师, CCF专业会员, 主要研究领域为无线传感器网络, 移动计算, 深度学习.

刘念伯(1975－), 男, 博士, 副研究员, 主要研究领域为无线传感器网络, 移动计算, 人工智能.