

基于深度强化学习的双模通信超帧时隙分配算法

冯俊豪^{1,2}, 詹学滋³, 颜宇帆⁴, 王朋姣⁴, 陈智雄^{3,4}

(1. 南方电网科学研究院, 广州 510663; 2. 广东省电网智能量测与先进计量企业重点实验室, 广州 510603; 3. 华北电力大学电子与通信工程系, 河北 保定 071003; 4. 电力物联智慧化技术河北省工程研究中心, 河北 保定 071003)

摘要: 电力线与无线双模通信混合组网可实现优势互补, 在电力物联网、智能家居等领域具备重要的应用前景。为提升双模通信MAC算法的接入灵活性与网络资源利用率, 提出了一种基于深度强化学习的双模网关超帧时隙资源分配算法。首先构建了双模通信网关与终端的交互模型, 梳理了超帧结构及每个阶段的具体执行步骤; 接着定义了机器学习应用的关键奖励函数、状态空间与动作空间, 由网关对超帧接入的吞吐量等参数进行监测统计, 通过迭代学习与训练得到优化后的超帧时隙数、占比等参数。最后通过仿真验证了该算法的有效性与可靠性, 将所提算法与固定超帧结构、无超帧结构等算法进行对比, 分析了不同算法下网络吞吐量、平均时延、丢包率的表现以及关键参数对系统性能的影响规律。仿真结果表明, 所提出的算法可有效提升系统吞吐量、时延等方面的性能, 能够实现资源的灵活高效分配。

关键词: 双模通信; 双深度Q网络; MAC层接入; 超帧结构; 时隙分配

Super Frame Time Slot Allocation Algorithm for Dual-Mode Communication Based on Deep-Reinforcement Learning

FENG Junhao^{1,2}, ZHAN Xuezi³, YAN Yufan⁴, WANG Pengjiao⁴, CHEN Zhixiong^{3,4}

(1. Electric Power Research Institute, CSG, Guangzhou 510663, China; 2. Guangdong Provincial Key Laboratory of Intelligent Measurement and Advanced Metering of Power Grid, Guangzhou 510603, China; 3. Department of Electronics and Communication Engineering, North China Electric Power University, Baoding, Hebei 071003, China; 4. Power Internet of Things Smart Technology Engineering Research Center of Hebei Province, Baoding, Hebei 071003, China)

Abstract: The hybrid networking of power line and wireless dual-mode communication can achieve complementary advantages, demonstrating significant application prospects in fields such as the Internet of Things for power systems and smart homes. To enhance the access flexibility and network resource utilization of dual-mode communication MAC algorithms, a superframe time slot resource allocation algorithm for dual-mode gateways based on deep reinforcement learning is proposed. Firstly, an interaction model between the dual-mode communication gateway and terminals is established, detailing the superframe structure and the specific execution steps of each phase. Secondly, key reward functions, state spaces, and action spaces for machine learning applications are defined, with the gateway monitoring and collecting parameters such as superframe access throughput. Through iterative learning and training, optimized parameters including the number and proportion of superframe time slots are obtained. Finally, simulations are conducted to verify the effectiveness and reliability of the proposed algorithm. Comparisons are made with fixed superframe structure and non-superframe structure algorithms, analyzing network throughput, average delay, packet loss rate under different algorithms, as well as the influence patterns of key parameters on system performance. Simulation results indicate that the proposed algorithm can effectively improve system performance in terms of throughput and delay, enabling flexible and efficient resource allocation.

Key words: dual-mode communication; double deep Q-network; MAC layer access; super frame structure; time slot allocation

基金项目: 国家自然科学基金青年基金项目(61601182); 南方电网科学研究院科技项目(SEPRI-K23B060); 广东省科技计划项目(2021B1212050014)。

Foundation item: Supported by the Youth Funds of National Natural Science Foundation of China(61601182); the Science and Technology Projects of Electric Power Research Institute of China Southern Power Grid Co., Ltd.(SEPRI-K23B060); the Science and Technology Projects of Guangdong Province(2021B1212050014)。

0 引言

电力线通信(power line communication, PLC)和无线通信(wireless communication, WC)技术是智能电网和电力物联网的重要组成部分^[1-2]。电力线通信无需重新布线即实现业务接入^[3-5],但是存在噪声源多以及受负载影响大的特点。无线通信具备组网简单,不受负载变化影响等优点^[6],但是易受环境因素影响,不适合阴影和衰落影响较大的场景。因此,可利用电力线和无线双模通信实现双介质信道优势互补和灵活接入,综合提升网络整体性能。

目前,基于PLC和WC的混合通信系统研究聚焦于物理层和介质访问控制(media access control, MAC)层。文献[7]提出了一种混合PLC与自由空间光通信(free-space optical communication, FSO)的双跳混合通信网络,第一跳为PLC接入信道,第二跳为FSO传输信道,通过两跳混合网络延长传输距离。文献[8]结合了PLC与非正交多址接入技术,研究了对数正态分布衰落下PLC引入非正交多址接入的功率分配问题。文献[9]提出了适用于两跳混合中继的MAC层算法,综合考虑带冲突避免的载波侦听多路访问(carrier sense multiple access with collision avoidance, CSMA/CA)退避流程,建立跨层性能模型。上述研究大多采用CSMA/CA接入机制令终端节点竞争接入网络,而CSMA/CA竞争容易发生碰撞,导致丢包,因此可基于超帧(super frame, SF)结构,将时隙划分为无竞争和有竞争接入两个部分,终端节点可在非竞争时隙内进行时分多址(time division multiple access, TDMA)传输,减少竞争接入的机会,间接降低碰撞发生概率。

超帧在IEEE 802.15.4标准中得到应用,减少网络的碰撞概率的同时保证网络的可扩展性。文献[10]在电力线与无线双模通信网络中使用了CSMA/CA-TDMA超帧结构,在超帧的争用期内申请非争用期时隙,节点通过非争用期传输数据包。文献[11]提出的算法根据终端节点的休眠-工作特性,将终端节点划分为唤醒组和睡眠组,每个超帧只允许一个组的终端节点参与竞争,有效地提高了信道利用率,降低碰撞概率。文献[12]提出的方法通过智能算法控制终端节点在超帧争用期或者非争用期内访问信道,有效提高了系统性能,但是需要

大量的终端节点信息才能进行学习,算法效率较低。上述文献的研究都是从终端节点的访问角度出发开展的,但由于超帧中固定的时隙结构使得资源分配并不灵活。对此,可采用深度强化学习(deep reinforcement learning, DRL)设计灵活分配时隙资源的超帧接入算法,提高时隙资源分配的灵活性。

作为DRL的经典算法,深度Q网络(deep Q-network, DQN)^[13-14]已被应用于时频分配^[15-16]、导频^[17]、移动路径规划^[18]、分时节能^[19]等通信技术中。文献[20]通过DRL调整CSMA/CA争用窗口控制终端节点的接入频次,达到避免网络拥塞,减小碰撞概率的目的。文献[21]提出适用于非均匀步长情况下的改进深度强化学习多址接入(deep-reinforcement learning multiple access, DLMA)协议,且引入 α -公平效用函数避免智能体连续占用信道的不公平现象。文献[22]结合水声通信中时延较大的特点,设计了等待经验存储机制避免DQN训练的局部收敛现象。

从上述研究不难看出,DRL已在无线通信领域得到广泛应用,但针对PLC MAC层智能接入算法的相关研究仍处于起步阶段^[23],将DRL应用于双模通信的算法研究设计更是鲜有开展。DRL能够实现MAC层资源的灵活分配,有效降低算法复杂度。基于此,本文提出一种基于深度强化学习的电力线与无线双模通信网关超帧时隙分配算法(double-mode communication superframe slot allocation based on DRL, DCSSA-D),网关可通过DRL对超帧状态和历史信息进行动态学习,进而调整争用期与非争用期的大小,使全网吞吐量达到最优。为避免神经网络在学习过程中出现过估计问题,本文采用双深度Q网络(double deep Q-network, DDQN)替代传统DQN架构,在处理时间相近的条件下,提升了网关的决策精度。

1 网络模型与超帧结构

本节首先针对建立的电力线和无线双模通信网络模型,结合双模网关的功能定义,给出终端节点接入流程。接着定义双模接入的超帧结构以及超帧各个阶段的具体交互细节,提出面向网关的超帧时隙分配方法和具体工作过程。

1.1 电力线与无线双模通信网络模型

系统模型如图1所示。考虑基于电力线和无线

双模通信和综合网关的星形网络拓扑,其中中心协调网关G负责协调组网和网络资源分配, N 个双模通信终端节点 $I_i(i=1, 2, \dots, N)$ 可通过电力线和无线双媒质信道与网关G进行并行通信。显然,双模通信的使用提升了网络接入的灵活性,也为并行分集和复用等算法提供了技术支持。

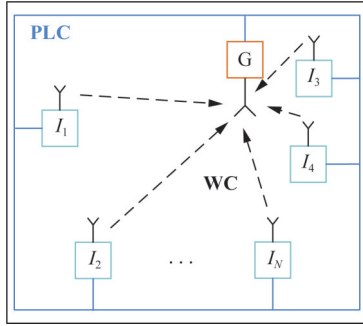


图1 双模通信网络模型

Fig. 1 Dual-mode communication network model

1.2 超帧结构

由于 PLC 和无线的 MAC 层均可定义基于 TDMA 和 CSMA 两种机制的超帧结构^[10],根据业务性能要求,进行灵活资源分配。因此,双模通信采用图2所示的超帧结构。为了聚焦算法设计,双模通信采用对称的双介质超帧结构,即无线信道和电力线信道使用相同的超帧结构。

超帧的基本单元为迷你时隙(长度为 T_{slot}),整体包括信标帧 T_{beacon} 、非争用时期(contention free period, CFP)和争用时期(contention period, CP)3个部分。信标帧为区分两个超帧的标志,网关G在信标帧内向全网终端节点I广播超帧的结构信息。

节点在争用期内通过 CSMA/CA 算法竞争接入,在非争用期内使用分配的 TDMA 时隙分时接入。争用期内,PLC 采用 IEEE 1901 标准^[24]的 CSMA 算法,无线信道采用 IEEE 802.11 标准^[25]的 CSMA 算法。

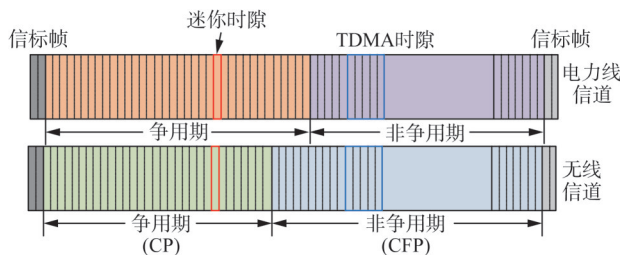


图2 超帧结构

Fig. 2 Superframe structure

在超帧的 CP 阶段,网关接收数据包并统计全网节点数据包的传输情况;在超帧的 CFP 阶段,节点按轮询表依次占用 TDMA 时隙,网关依次接收,并统计 CFP 阶段接收的数据包总数。网关将 CP 和 CFP 阶段接收的数据包数量与先前记录的决策信息存储,并以历史吞吐量、TDMA 时隙数量信息为状态参数,在下一轮超帧中输入 DDQN 中。在不断的决策训练迭代中,获得最优决策。

已知 CP、CFP 和信标帧长度分别为 T_{CP} 、 T_{CFP} 和 T_{beacon} ,则超帧总长度 T_{sf} 为:

$$T_{\text{sf}} = T_{\text{beacon}} + T_{\text{CP}} + T_{\text{CFP}} \quad (1)$$

已知 CP 和 CFP 占用 k 个和 $(M-k)$ 个迷你时隙,即 $T_{\text{CP}} = kT_{\text{slot}}$, $T_{\text{CFP}} = (M-k)T_{\text{slot}}$ 。为了便于管理和分配,将 CFP 中 l 个迷你时隙组成一个 TDMA 时隙(长度为 T_{TDMA}),若 CFP 由 L 个 TDMA 时隙组成,则满足以下关系: $T_{\text{TDMA}} = lT_{\text{slot}}$, $T_{\text{CFP}} = LT_{\text{TDMA}}$ 。

节点通过接收网关广播的超帧信息与 ACK 消息实现高效网络接入。在争用期内若网关成功接收到数据包,会向全网广播该数据包的 ACK 消息,标记本次传输成功并记录传输结果;若网关未成功接收数据包,则向全网广播数据包传输失败消息。在非争用期内所有节点完成传输后,网关会向全网广播 ACK 消息,并统计 CFP 阶段成功传输的数据包数量。为提升网络整体性能,网关将依托采集到的网络信息,通过深度强化学习 DRL 确定接近最优的超帧结构,给出 CP 与 CFP 阶段的占比方案以及 TDMA 时隙分配数量。

2 基于 DRL 的网关超帧时隙分配算法

以 DDQN 等为代表的深度强化学习(deep reinforcement learning, DRL)算法,可用于处理具有高维状态和动作空间的复杂任务,被广泛应用于通信资源分配中。它利用深度神经网络来逼近强化学习中的策略函数和价值函数,从而提高学习能力和决策效率^[13]。双模通信网关采用 DRL 算法分配超帧时隙,本节将先介绍算法的具体步骤,重点给出 DRL 算法的输入输出参数、环境状态、动作、奖励以及 Q 值的定义,分析并总结决策模块和神经网络训练模块的工作原理。

2.1 DCSSA-D 算法流程

由于 DDQN 相比于 DQN,通过采用两个独立的网络减少了 Q 值的过度估计现象,并且提高离散

状态下的决策稳定性,因此DCSSA-D中引入DDQN来代替传统的DQN架构。

网关通过在CP和CFP阶段侦听并存储历史信息来获得当前网络环境状态,具体工作过程如图3所示。 Q 值定义为网关G在状态为 s_t 时采取动作 a_t 得到的累积折扣奖励期望值。

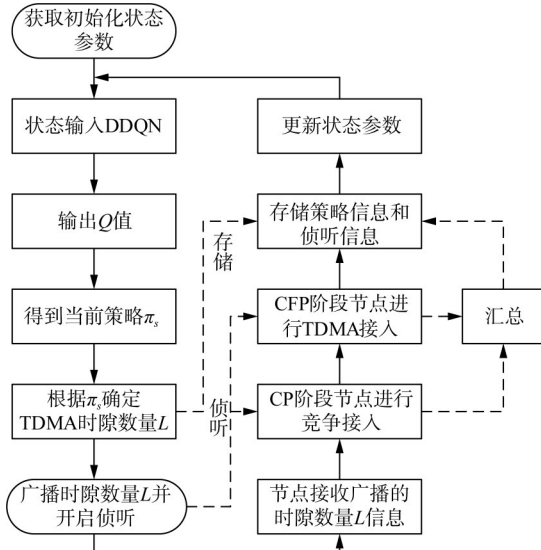


图3 网关工作过程

Fig. 3 Working process of the gateway

算法的主要流程如下。

- 1) 算法初始化: 超帧结构参数、网络吞吐量等历史数据;
- 2) 网关将当前状态参数(先前 W 个超帧的吞吐量和TDMA时隙数量)输入DDQN;
- 3) DDQN输出 Q 值,并由 ε 贪婪策略作出决策 π_s ;
- 4) 根据 π_s 确定当前超帧TDMA时隙数量和使用次序,广播给各个节点;
- 5) CP阶段,网关接收并及时统计数据包数量等信息;
- 6) CFP阶段,网关根据决策次序依次接收各节点在TDMA时隙传输的数据包;
- 7) 网关统计超帧吞吐量,获得当前奖励值;
- 8) 根据超帧吞吐量和TDMA时隙数量更新状态参数(即下一次迭代的状态参数),生成经验放入经验池中;
- 9) 开展经验重放,更新DDQN;
- 10) 进入下一轮迭代。

2.2 DDQN定义和迭代计算

DRL过程示意如图4所示,DCSSA-D算法结构包括训练模块和决策模块两部分。训练模块负责生成 Q 值和神经网络迭代更新,决策模块负责根据训练模块给出的 Q 值形成最优决策。

在训练阶段,网关G需要对经验池中的经验 $E_t = [s_t, a_t, r_t, s_{t+1}]$ 进行抽样,抽取部分经验进行经验重放。网关G通过经验重放开展学习。采用 ε 贪婪策略权衡网关G进行随机决策和最优决策的倾向,使网关G能够遍历尽量多的网络环境状态,获得更多样化的奖励。除此之外,为避免网关G的DRL算法出现 Q 值过高估计而导致决策错误,本文采用DDQN架构构建DRL学习模型。

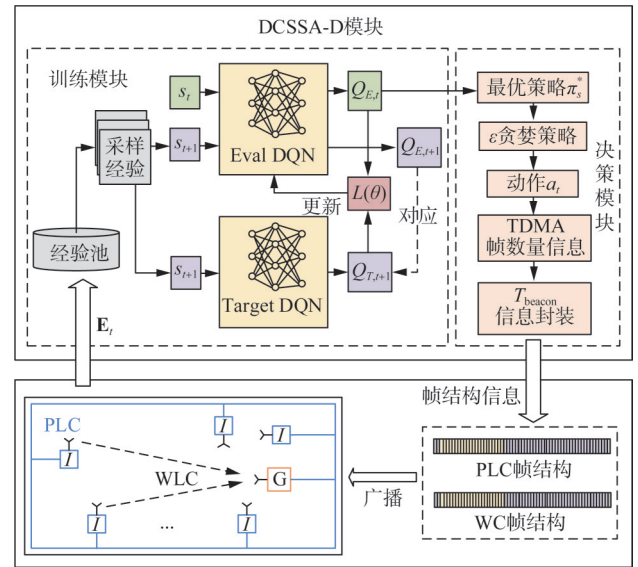


图4 DDQN流程

Fig. 4 DDQN process

1) 动作

设网关在超帧 t 时做出的动作为 $a_t = L_t$,其中 L_t^{PLC} 和 L_t^{WC} 分别为电力线信道和无线信道一个超帧中TDMA时隙数量,动作空间为 A , $a_t \in A$ 。

由于超帧长度 T_{sf} 是固定的,因此,若网关选择信道 c 的超帧中TDMA时隙数量为 $L_t^{(c)}$,那么该信道争用期CP长度为:

$$T_{\text{CP},t}^{(c)} = T_{\text{sf}} - L_t^{(c)} T_{\text{TDMA}} - T_{\text{beacon}} \quad (2)$$

当信道数据流量较小时可减少 $L_t^{(c)}$ 或增加CP时隙数,节点可采用竞争接入的方式传输更多数据包。若数据流量较大,可增加 $L_t^{(c)}$ 大小,更多节点加入TDMA轮询分配,避免碰撞导致吞吐量性能

下降。

2) 状态空间

本文以网络吞吐量 μ 和动作 a_t 相结合的形式来构建网络状态 s_t 。在设定环境状态时还考虑了历史信息，即以往的超帧结构和网络流量数据。

设在超帧 t 中网络的吞吐量 μ_t 为：

$$\mu_t = \frac{\sum_{i=1}^N \Pi_{t,i}}{T_{s_t}} \quad (3)$$

式中 $\Pi_{t,i}$ 为终端节点 i 在超帧 t 中传输的数据量，bit，则当前超帧 t 时的网络状态 o_t 为：

$$o_t = [\mu_t, a_t] \quad (4)$$

则网络环境状态 s_t 为先前 W 个超帧 o_t 的集合，即：

$$s_t = \{o_{t-W}, \dots, o_{t-1}\} \quad (5)$$

式中 o_{t-W} 和 o_{t-1} 分别为超帧 $t-W$ 和 $t-1$ 时的网络状态。

3) 奖励函数

若在一个超帧内传输的数据包数量越多，信道资源利用率越高，网关将得到更多的奖励。为简化算法，本文考虑超帧 t 内传输的数据包数量以及发生碰撞的次数生成奖励值，即 $r_t = \delta_t$ 。其中， δ_t 为两条信道在超帧 t 内传输的数据包数量。

4) Q 值计算和神经网络训练

设在第 t 个超帧中DQN获得的累积折扣奖励 R_t 为：

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \quad (6)$$

式中： γ 为累积折扣奖励因子， $\gamma \in [0, 1]$ ； k 为累积奖励中的步数； r_{t+k+1} 为 $t+k+1$ 个超帧的即时奖励值。由此，可得当网络环境状态为 s_t ，网关 G 采取动作 a_t 时DQN输出的 Q 值为：

$$Q(s_t, a_t; \theta) \triangleq E[R_t | s_t, a_t, \theta] \quad (7)$$

式中： $Q(s_t, a_t; \theta)$ 为网关 G 在状态为 s_t 时采取动作 a_t 得到的累积折扣奖励期望值，其中 θ 为神经网络性能参数集； $E[\cdot]$ 为数学期望函数； R_t 为第 t 个超帧对应的累积折扣奖励。

为了避免过拟合且及时更新状态，DDQN将下一时隙状态 s_{t+1} 输入到Eval网络中，然后在Target网络中找出相同决策产生的对应 Q 值，利用这两个 Q 值对神经网络进行参数更新。设DDQN在 $t+1$ 时刻采取动作 a_t 时Eval网络的 Q 值 $Q_{E,t+1}$ 和Target网

络输出的 Q 值 $Q_{T,t+1}$ 分别为：

$$\begin{cases} Q_{E,t+1} = Q(s_{t+1}, a_t; \theta_{E,t}) \\ Q_{T,t+1} = Q(s_{t+1}, a_t; \theta_{T,t}) \end{cases} \quad (8)$$

式中 $\theta_{E,t}$ 、 $\theta_{T,t}$ 分别为Eval网络和Target网络在超帧 t 时的神经网络性能参数集。由此可得 Q 值的更新函数 $Q_{E,t+1}$ 为：

$$Q_{E,t+1} \leftarrow R_{t+1} + \beta Q(s_{t+1}, \arg \max_{a_t \in A} Q_{E,t+1}; \theta_{T,t+1}) \quad (9)$$

式中： R_{t+1} 为第 $t+1$ 个超帧中DQN获得的累积折扣奖励； β 为折扣因子； A 为动作空间。

更新过程中损失函数 $L(\theta)$ 为：

$$L(\theta) = \frac{1}{N_E} \sum_{e \in E} (R_{t+1} + \beta Q_{T,t+1} - Q_{E,t})^2 \quad (10)$$

式中： N_E 为经验池抽样数量； e 为单条经验样本； E 为经验池中抽样得到的经验样本集合。

5) 最优策略

设 $Q_{E,t}^*(s_t, a_t; \theta_{E,t})$ 为网关 G 在网络状态为 s 时采取策略所能达到的最优的 Q 值， π_s^* 为网关 G 在网络状态为 s 时所产生的最优策略，则有：

$$\pi_s^* = \arg \max_{a_t \in A} Q_{E,t}^*(s_t, a_t; \theta_{E,t}) \quad (11)$$

式中： A 为网关可选动作集合； a_t 为网关在超帧 t 采取的动作。

在DRL算法进行学习初期，需要累积足够多的经验，从而在后续进行经验重放时不会因为无法遍历足够的状态而导致决策错误。因此，网关在决策过程中采用 ε 贪婪策略，即

$$a_t = \begin{cases} \text{rand}(A), & \text{概率为 } \varepsilon \\ \pi_s^*, & \text{概率为 } 1 - \varepsilon \end{cases} \quad (12)$$

式中： π_s^* 为网关 G 在网络状态 s 下的最优策略； ε 为网关随机决策的概率， $\varepsilon \in (0, 1]$ ，即在一次决策过程中网关 G 有 ε 的概率随机进行决策，即从确认集 A 中随机选取一个动作，同时有 $1 - \varepsilon$ 的概率做出最优决策。随着迭代次数的增加， ε 值将迭代减小，网关最终获得相对稳定的最优策略。

3 仿真分析

本节从吞吐量、时延和碰撞概率3个角度对比本文算法与各类接入算法和分配算法的性能。除此之外，本节还对比了使用DDQN和DQN两种DRL算法对DCSSA-D算法性能的影响。仿真在Python 3.8环境下进行，通过Tensorflow2.0搭建DDQN架

构, 使用Pycharm解释器构建仿真算法, 所得数据通过MATLAB进行处理。

3.1 网络设置和对比算法

本文以吞吐量 μ 、接入时延和碰撞概率作为反映算法性能的指标。节点在传输过程中系统参数设置如表1所示。

表1 网络参数设置

Tab. 1 Network parameter settings

名称	数值
$E[P]$ /bits	12 480
ACK值/bits	240
传输速率/(Mbit·s ⁻¹)	6
时隙大小/ μ s	10
SIFS值/ μ s	16
DIFS值/ μ s	34

假设网络中只有发生数据包碰撞才会导致丢包, 所有的节点都处于饱和发送状态。每个节点都具有相同的优先级, 其中电力线单模节点竞争窗口值 $CW_1 = [8, 16, 32, 64]$, 延迟计数器值 $DC = [0, 1, 3, 15]$; 无线单模节点的竞争窗口值 $CW_2 = [32, 64, 128, 256]$ 。表中 $E[P]$ 为节点发送的数据包长度, 所有节点发送的数据包都是等长的; ACK为当节点成功传送后网关向全网广播的确认信息; 每个时隙大小为 10μ s, 分布式帧间间隙(distributed inter-frame spacing, DIFS)大小为 34μ s, 短帧间间隔(short inter-frame spacing, SIFS)大小为 16μ s。

本节讨论了电力线和无线双模通信网络模型在使用DCSSA-D、复用接入^[24]、并行接入^[9]算法对网络性能的影响。对于复用接入算法, 课题考虑了两种情况: 1)不使用超帧, 节点通过两条信道竞争接入; 2)使用超帧, 固定时隙, 节点在CP阶段用来申请CFP中TDMA时隙资源。对于并行算法, 本文考虑了文献[23]中的情况, 即使用超帧且固定时隙。本文考虑超帧结构具体参数如表2所示。

3.2 仿真设置

参考文献[11]并结合仿真调试, 本文的DRL参数设置如表3所示。

考虑到本文的状态 s_t 设置为先前 W 个时隙的 o_t 的集合, 可以看作是时间序列。因此, 擅长处理长期记忆数据的循环神经网络(recurrent neural

表2 超帧结构参数

Tab. 2 Parameters of super frame structure

参数	取值
T_{sf}/μ s	52 000
T_{beacon}/μ s	50
T_{CF}/μ s	31 200
T_{CFP}/μ s	20 750
L	10
T_{slot}/μ s	10

表3 DDQN超参数设置

Tab. 3 DDQN hyperparameters

超参数	取值
状态历史长度 W	20
ϵ 取值	$[0.005, 1]$
λ	0.995
折扣因子 γ	0.9
经验池大小	1 000
抽样经验池大小 N_E	128
Target网络更新频率	20

network, RNN)为本算法模型中DDQN神经网络的最优选^[26]。具体来说, 在隐藏层中加入一层长短期记忆网络(long short-term memory, LSTM)。

如图5所示, 神经网络结构中包括一个输入层、两个隐藏层和一个输出层。其中一层隐藏层为LSTM, 在LSTM后面的隐藏层为前馈神经网络(feedforward neural network, FNN), 相比传统DDQN架构能够同时捕捉时间序列中的长期依赖性和复杂状态的非线性特征表达, 提升决策过程的准确性。神经网络的输出为在当前状态下网关采取所有动作所产生的 Q 值。在神经网络的超参数设置上, 隐藏层上每层具有64个神经元, 每个神经元的激活函数为ReLU。

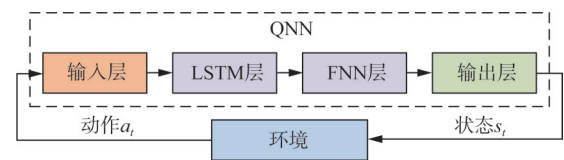


图5 神经网络结构示意图

Fig. 5 Schematic diagram of neural network structure

选择较大的状态历史长度 W 来覆盖足够多的历史状态, 以学习更多潜在的MAC接入行为。具体来说, 设置 $W = 20$, 即网关在第 t 个超帧进行迭代时, 输入的环境状态 s_t 覆盖了20个历史超帧长度。

对于 ε 贪婪策略的参数设置, 本文采用逐步迭代减小的方式更新 ε 值。即 $\varepsilon_{i+1} \leftarrow \lambda \varepsilon_i$, 其中 λ 为每次迭代的迭代因子。本文设置 ε 初值为 1, 终值为 0.005, 迭代因子 λ 为 0.995。

为满足神经网络训练要求, 经验重放所使用的经验池大小为 1 000, 进行经验重放时, 从经验池中随机抽取 128 个经验。为进一步提高神经网络的收敛速度和缩短训练时间, 本文使用 RMSProp 优化器神经网络进行小批量梯度下降。DDQN 中的 Target 网络每隔 20 次迭代就复制当前的 Eval 网络, 完成神经网络参数更新。

参数选择后得到神经网络模型的 Loss 函数如图 6 所示, Loss 函数仿真过程中, 网络中节点数量 $N = 40$ 。本文设置当经过 200 个超帧, 即 $t = 200$ 时开始训练神经网络。

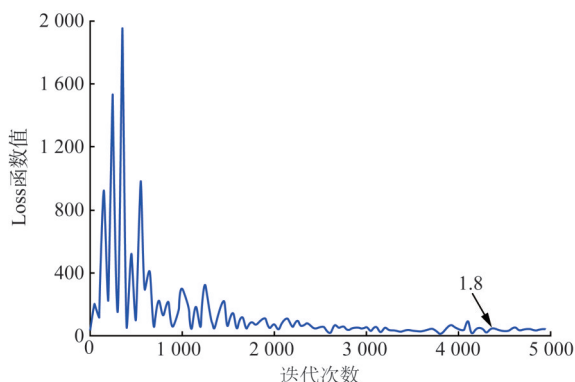


图 6 损失函数

Fig. 6 Loss function

从图 6 中可以发现, 当迭代次数小于 2 000 次之前, Loss 函数值具有较大波动, 这是因为 DCSSA-D 网关未能收集足够多的经验, 通过不断试错获得尽量多的结果。当迭代次数大于 2 000 左右时, Loss 函数值趋于稳定, 在迭代次数大于 4 000 之后, Loss 值稳定在 1.8 左右, 表明此时神经网络已经找到了最优决策, 算法收敛。

3.3 仿真分析

本节讨论了使用 DCSSA-D 算法对网络性能的影响, 并与非智能算法进行对比。对比算法总体可分为使用超帧结构和不使用超帧结构两种类型。对于使用超帧结构的对比算法, 引入了文献[11-12]的算法。其中, 文献[11]的算法中, 节点通过 CP 申请占用 CFP 中的 TDMA 时隙, 利用 TDMA 时隙传输数据包(超帧申请); 文献[12]中节点在 CP 中

通过 CSMA/CA 算法竞争接入, 在 CFP 中通过 TDMA 算法接入(超帧传输)。对于不使用超帧结构的算法, 节点通过文献[23]提出的算法复用竞争接入(CSMA/CA)。

1) 网络吞吐量性能

图 7 为网络的总吞吐量随着网络中节点数量的变化情况。图中对并行接入超帧、复用接入超帧、复用接入 CSMA/CA 和采用 DQN 架构的 DCSSA-D 进行了对比。

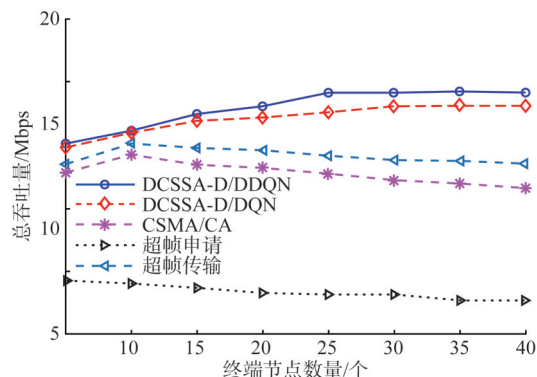


图 7 总吞吐量

Fig. 7 Total throughputs

从图 7 中可以看出, DCSSA-D 算法对比其它算法在吞吐量性能上具有一定的优势。DCSSA-D 算法能够灵活决定超帧数量, 从而让全网的吞吐量指标始终保持在最优的水平, 因此在节点数量增加的情况下, 总吞吐量性能下降较不明显。

2) 平均接入时延性能

图 8 为网络中节点平均接入时延随着终端节点数量的变化情况。

随着终端节点数量的增加, 节点的平均时延也

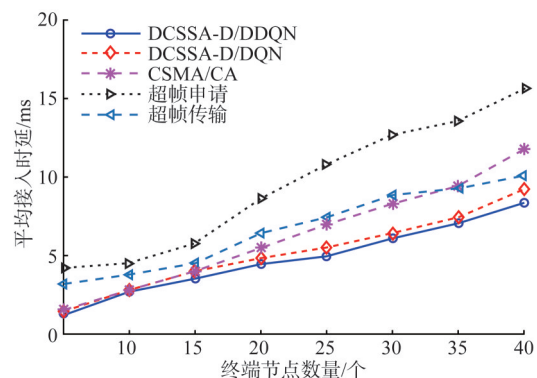


图 8 平均接入时延

Fig. 8 Average access latencies

随之增加。对比图中其他算法,本文提出的DCSSA-D算法具有更低的接入时延。该算法根据网络状态做出决策控制超帧中CFP的大小,节点数量较少时流量较小,采用CSMA/CA竞争发送时发生碰撞的概率较低;当网络中节点数量较多时,网关利用DCSSA-D算法能够安排更多的TDMA时隙,减小CP的长度,使得更多的终端节点以TDMA接入的方式接入信道,尽量避免信道中发生碰撞,使得丢包率相对其他算法而言更小。

3) 丢包率性能

图9为网络总丢包率随着终端节点数量变化的变化情况。超帧申请算法中丢包率较高的原因是由于TDMA时隙资源占满而导致丢包。

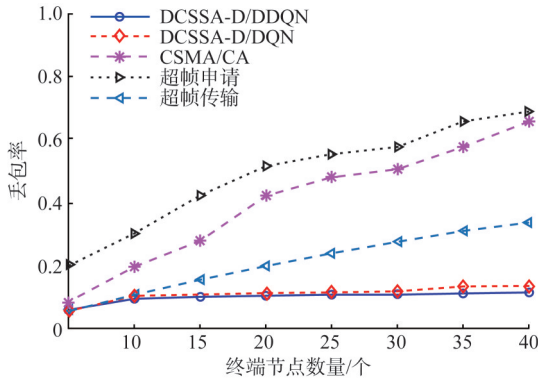


图9 丢包率

Fig. 9 Packet loss rates

由图9可知,采用DCSSA-D算法的网络丢包率相较于其他算法有明显降低。由于DCSSA-D算法在终端节点数量增加时,为CFP分配更多的时隙资源,使得超帧中TDMA时隙数量增加,CFP可承载更多的数据包,以减少碰撞概率。

4) 平均TDMA时隙数量

图10为在不同终端节点数量条件下,执行DCSSA-D算法时超帧中TDMA时隙数量。

当节点数量较小时,为了避免TDMA时隙资源浪费,网关调控较少数量的TDMA时隙。当网络中终端节点数量增加时,由于CP中采用CSMA/CA竞争接入产生的碰撞次数增加,导致吞吐量减小,此时网关增加TDMA时隙数量,减小网络中碰撞的发生,尽可能保证吞吐量大小。

争用窗口对平均TDMA时隙数量具有一定的影响,且这种影响在节点数量较低时更为明显。在节点数量较低时,如果节点具备较大的争用窗口,

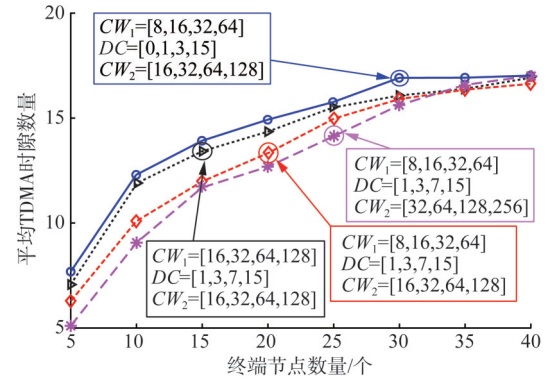


图10 平均TDMA时隙数量

Fig. 10 Average number of TDMA slots

CP中产生更多信道空闲。而当节点数量较大时,TDMA时隙资源一直处于饱和状态,因此终端节点数量较大时,争用窗口对TDMA时隙数量影响较小。无线争用窗口对TDMA时隙数量影响大于电力线争用窗口。这是因为IEEE 802.11的BEB机制中没有DC延迟计数器,因而无线信道中的流量更高,因而对TDMA时隙数量具有更大的影响。

5) 平均奖励

图11为当网络终端节点数量增加时,DCSSA-D网关获得的奖励情况。平均奖励为在仿真过程中每一次超帧中成功传输的次数取平均值。DCSSA-D算法与复用超帧算法相比,能够灵活调控CFP的大小获得更大的奖励收益,保障系统吞吐量性能。

当DCSSA-D算法采用DDQN架构时,其获得的平均奖励比使用DQN架构更高。采用DQN架构时,奖励值会有不稳定现象的发生。这是因为DQN算法在迭代时会出现Q值过估计现象,使得网关容易产生非最优决策,因此使用DQN架构的DCSSA-D算法比使用DDQN架构的平均奖励更低。

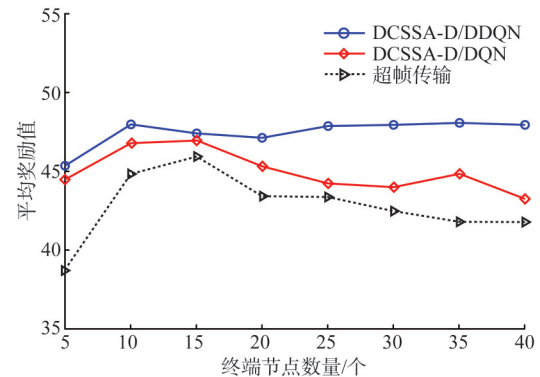


图11 平均奖励

Fig. 11 Average reward

4 结语

本文提出了一种基于深度强化学习(DRL)的电力线与无线双模通信网关超帧时隙分配算法(DCSSA-D)。该算法通过双重深度Q网络(DDQN)实现超帧时隙的优化分配,得到准最优的超帧竞争接入期(CP)、无竞争期(CFP)长度,以及时分多址(TDMA)的迷你时隙数量,可在不同终端数量条件下优化提升网络吞吐性能。

与现有其他算法相比,DCSSA-D算法不仅能够提升网络吞吐量,还可降低终端接入的平均时延、减少碰撞概率。

后续可围绕两个方向开展研究:一是研究双模异构通信网络中基于Actor-Critic等DRL算法的自适应接入机制,通过负载均衡避免单一节点占用过多资源;二是研究单模节点与双模节点共存场景下基于DRL的网关时隙分配算法。

参考文献

- [1] DIB DE M B A L, FERNANDES V, DE L. FILOMENO M, et al. Hybrid PLC/wireless communication for smart grids and internet of things applications [J]. IEEE Internet of Things Journal, 2018, 5(2): 655 – 667.
- [2] 代洲,张全玲,凌琴,等.面向电力线通信融合双模态索引和时间分集的差分混沌移位方案[J].电力系统保护与控制, 2025, 53(15): 135 – 145.
DAI Zhou, ZHANG Quanling, LING Qin, et al. Dual mode index and time diversity based DCSK for power line communications[J]. Power System Protection and Control, 2025, 53(15): 135 – 145.
- [3] 王清,荆臻,李琼琮,等.宽带电力线载波通信物理层资源分配新算法[J].电测与仪表, 2021, 58(7): 124 – 129.
WANG Qing, JING Zhen, LI Congcong, et al. A novel algorithm of physical layer resource allocation for broadband power line carrier communication[J]. Electrical Measurement & Instrumentation, 2021, 58(7): 124 – 129.
- [4] YOON S G, JANG S, KIM Y H, et al. Opportunistic routing for smart grid with power line communication access networks [J]. IEEE Transactions on Smart Grid, 2014, 5(1): 303 – 311.
- [5] PAPADOPOULOS T A, KALOUDAS C G, CHRISTOS A I, et al. Application of narrowband power-line communication in medium-voltage smart distribution grids[J]. IEEE Transactions on Power Delivery, 2013, 28(2): 981 – 988.
- [6] 杨璐,吴清亮.无线接入网络自适应公平调度算法[J].通信学报, 2012, 33(1): 102 – 106, 113.
YANG Lu, WU Qingliang. Self-adaptive fair scheduling algorithm in wireless network[J]. Journal on Communications, 2012, 33(1): 102 – 106, 113.
- [7] 陈生海,言小琴,黎赛,等.混合双跳PLC-FSO通信系统的性能分析[J].通信学报, 2021, 42(10): 243 – 250.
CHEN Shenghai, YAN Xiaoqin, LI Sai, et al. Performance analysis of dual-hop PLC-FSO communication system [J]. Journal on Communication, 2021, 42(10): 243 – 250.
- [8] 曹旺斌,唐宏凯,谢志远,等.下行NOMA-PLC系统最优功率分配方法研究[J].中国电机工程学报, 2025, 45(3): 834 – 846.
CAO Wangbin, TANG Hongkai, XIE Zhiyuan, et al. Research on optimal power allocation scheme in downlink NOMA-PLC system [J]. Proceedings of the CSEE, 2025, 45(3): 834 – 846.
- [9] 黄鑫,卞宝银,杨鸿珍,等.无线和电力线混合通信的MAC层接入算法设计和性能分析[J].电力信息与通信技术, 2024, 22(1): 77 – 85.
HUANG Xin, BIAN Baoyin, YANG Hongzhen, et al. Design and performance analysis of MAC layer access algorithms for hybrid communication of wireless and power line [J]. Electric Power Information and Communication Technology, 2024, 22(1): 77 – 85.
- [10] CHEN Z, ZHI L, ZHAO X, et al. Parallel access scheme for wireless and power line dual-connection communication [J]. IEEE Transactions on Vehicular Technology, 2022, 71(7): 7997 – 8001.
- [11] AL-JANABI T A, AL-RAWESHIDY H S. An energy efficient hybrid MAC protocol with dynamic sleep-based scheduling for high density IoT networks[J]. IEEE Internet of Things Journal, 2019, 6(2): 2273 – 2287.
- [12] SHRESTHA B, HOSSAIN E, CHOI K W. Distributed and centralized hybrid CSMA/CA-TDMA schemes for single-hop wireless networks [J]. IEEE Transactions on Wireless Communications, 2014, 13(7): 4050 – 4065.
- [13] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518(7540): 529 – 533.
- [14] NGUYEN T T, NGUYEN N D, NAHAVANDI S. Deep reinforcement learning for multiagent systems: a review of challenges, solutions, and applications [J]. IEEE Transactions on Cybernetics, 2020, 50(9): 3826 – 3839.
- [15] 赵晓荣.基于深度强化学习的异构网络资源分配[D].济南:山东大学, 2023.
- [16] 姜华,杨家伟,黄巍,等.基于深度强化学习的D2D辅助MEC网络资源分配算法[J].电力信息与通信技术, 2023, 21(7): 51 – 58.
JIANG Hua, YANG Jiawei, HUANG Wei, et al. A D2D-assisted MEC network resource allocation algorithm based on deep reinforcement learning [J]. Electric Power Information and Communication Technology, 2023, 51(7): 51 – 58.
- [17] 刘乔寿,周雄,刘爽,等.基于深度强化学习的OFDM自适应导频设计[J].通信学报, 2023, 44(9): 104 – 114.
LIU Qiaoshou, ZHOU Xiong, LIU Shuang, et al. Adaptive pilot design for OFDM based on deep reinforcement learning [J]. Journal on Communications, 2023, 44(9): 104 – 114.
- [18] 赵玉新,杜登辉,成小会,等.基于强化学习的海洋移动观测网络观测路径规划方法[J].智能系统学报, 2022, 17(1): 192 – 200.
ZHAO Yuxin, DU Denghui, CHENG Xiaohui, et al. Path planning for mobile ocean observation network based on reinforcement learning [J]. CAAI Transactions on Intelligent Systems, 2022, 17(1): 192 – 200.

(下转第 121 页 Continued on Page 121)