

基于智能分层切片技术的数字孪生传感信息同步策略

唐伦 李质莹* 文雯 成章超 陈前斌

(重庆邮电大学通信与信息工程学院 重庆 400065)

(移动通信技术重庆市重点实验室 重庆 400065)

摘要: 针对传感数据在无线接入网(RAN)中传输的不可靠性与不及时性造成数字孪生(DTs)同步信息的不精确问题, 该文提出一种基于智能分层切片技术的DTs传感信息同步策略。该策略在双时间尺度下, 以最大化传感信息满意度和最小化切片重配置及DTs同步成本为目标, 联合优化切片无线资源配置以及DTs传感信息同步问题。首先, 在大时间尺度, 利用网络切片为有着不同服务质量(QoS)的DTs提供隔离以及解决部署问题; 在小时间尺度, 通过更加灵活的无线资源分配来提高DTs传感信息同步任务对动态环境的适应性, 进一步提高通信性能, 建立更逼近于物理实体的DTs。其次, 为了求解不同时间尺度的优化问题, 该文提出一种双层深度强化学习(DRL)框架实现高效的网络资源交互, 其中下层控制算法利用优先经验放回(PER)机制加快收敛速度。最后, 仿真结果验证了所提策略的有效性。

关键词: 数字孪生; 网络切片; 深度强化学习; 状态估计; 资源分配

中图分类号: TN929.5

文献标识码: A

文章编号: 1009-5896(2022)YU-0001-10

DOI: 10.11999/JEIT230984

Digital Twin Sensing Information Synchronization Strategy Based on Intelligent Hierarchical Slicing Technique

TANG Lun LI Zhixuan WEN Wen CHENG Zhangchao CHEN Qianbin

(School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

(Chongqing Key Laboratory of Mobile Communications Technology, Chongqing 400065, China)

Abstract: In order to mitigate the problem of inaccurate synchronization sensory information in Digital Twins (DTs) caused by unreliable and delayed transmission in Radio Access Networks (RAN), a sensory information synchronization strategy for DTs based on intelligent hierarchical slicing technology is proposed. The strategy aims to optimize the allocation of wireless resources for slicing and the synchronization of DTs' sensing information in dual time scales, with the goals of maximizing the satisfaction of sensing information and minimizing the costs associated with slicing reconfiguration and DTs' synchronization. Firstly, at large time scales, network slicing is employed to provide isolation for DTs with varying Quality of Service (QoS) and resolve deployment challenges; At small time scales, a more flexible wireless resource allocation is utilized to enhance the adaptability of DTs' sensory information synchronization to dynamic environments. Secondly, in order to optimize the synchronization of DTs' sensory information at different time scales, a two-layer Deep Reinforcement Learning (DRL) framework is introduced to facilitate efficient network resource interaction, and in the framework the lower-layer control algorithm incorporates the Prioritized Experience Replay (PER) mechanism to accelerate convergence speed. Finally, the effectiveness of the proposed strategy is validated through simulation results.

Key words: Digital Twin (DT); Network slice; Deep Reinforcement Learning (DRL); State estimation; Resource allocation

收稿日期: 2023-09-07; 改回日期: 2023-12-12; 网络出版: 2023-12-22

*通信作者: 李质莹 1002012709@qq.com

基金项目: 国家自然科学基金(62071078), 四川省科技计划(2021YFQ0053)

Foundation Items: The National Natural Science Foundation of China (62071078), Sichuan Science and Technology Program (2021YFQ0053)

1 引言

数字孪生(Digital Twins, DTs)作为6G网络的重要应用技术,引起了广泛的关注^[1-3]。DTs是物理世界在虚拟空间的数字化映射^[4],对实体的行为实时监控和同步,实现实体的精确数值表示,还具有模拟、预测以及全生命周期管理等功能,因此DTs能帮助6G网络提高网络性能、管理效率并减少运营成本^[5]。

现有很多文献集中研究如何利用DTs辅助网络进行资源优化,但其大多基于DTs已建立完善的前提^[6-8],如何让DTs精确地反映实体也是DTs与6G网络集成亟需解决的问题。传感数据是DTs建模的重要数据源,本文利用这些数据估计设备的状态,以间接得到一些无法直接测量得到的数据^[9],为DTs精准表达实体提供数值支持。然而,不合理的资源分配将导致信息估计的准确度下降^[10,11],难以满足DTs对于信息同步的服务质量(Quality of Service, QoS)要求。网络切片^[12-14]作为一种新兴技术,能够解决这一挑战。

网络切片能够将一个物理网络划分为多个虚拟网络^[12],针对不同业务提供定制化的资源分配方案。采用双时间尺度框架的分层网络切片能够根据长期和短期资源需求的变化趋势,进行不同粒度的资源分配^[13,14],故较于单时间尺度,双时间尺度的结构能为DTs提供更灵活的切片功能,提高无线资源的利用效率。

在无线接入网(Radio Access Network, RAN)中,传输的不可靠性与不及时性会降低DTs同步信息的精确度,为此,本文提出一种基于智能分层切片的数字孪生传感信息同步策略,来提高DTs同步信息的精确度。主要工作总结如下:

首先,本文设计一个DTs驱动的RAN切片框架。其次,本文利用分层网络切片给DTs估计信息同步提供定制化服务,以最大化传感信息满意度和最小化切片重配置及DTs同步成本为目标,构建了数学模型。具体来说,在大时间尺度,上层控制策略动态调整为DTs分配的频谱资源、DTs的放置和迁移决策。在上层资源分配的约束下,下层控制策略在小时间尺度内采用状态估计的均方误差(Mean Square Error, MSE)评估估计性能,灵活配置功率、带宽。最后,针对双时间尺度问题之间存在耦合关系以及混合动作空间较大的情况,本文提出分层深度强化学习来解决这些问题,其中,下层采用的优先经验回放的多智体深度确定性梯度(Prioritized Experience Replay Multi-Agent Deep Deterministic Policy Gradient, PER-MADDPG)算法利用

PER机制来加快收敛速度以及多智体的协作求解优化问题;在下层控制收敛情况下,上层控制使用双深度Q网络(Double Deep Q Network, DDQN)调整切片缩放规模与完成DTs的迁移及放置。

2 系统模型

2.1 DTs支持的无线接入网切片场景

DTs支持的RAN切片场景由两层、两个闭环组成,如图1所示。其中两层分别为物理层和DTs层,两个闭环分别为内闭环和外闭环。内闭环为DTs基础模块向DTs支持的切片管理模块发送数据,切片管理模块对DTs基础模块进行管理;外闭环为DTs层对物理层进行资源分配,物理层向DTs层发送反馈。

物理层由物联网传感设备、基站、边缘服务器等设备组成。DTs层包括两个模块:DTs基础模块和DTs支持的切片管理模块。DTs基础模块包括DTs数据仓和DTs模型;DTs数据仓通过对估计的数据(温度、湿度、姿态等)进行采集、存储,为DTs模型的建立提供数据支持。DTs模型分为建立DTs需要的软件模块以及映射到边缘服务器的DTs。DTs支持的切片管理模块对DTs基础模块进行维护及管理,该模块包括DTs模型管理、网络切片管理、以及DTs功能管理。

2.2 数字孪生同步模型

本文将传感设备的DTs建立在基站配备的边缘服务器中以实现低延迟的服务^[15],考虑到边缘存储能力有限,故采用云边混合部署的方式,在云端的移动边缘计算(Mobile Edge Computing, MEC)服务器中缓存创建DTs所需的软件模块。当迁移发生时,边缘处的DTs从云端下载软件模块,完成创建与实例化。

因此设备*i*的DT表示为 $DT^d(i) = \{L_i, \Delta L_i(t), S_i(t), \mathcal{V}_i\}$,其中, L_i 为设备历史数据,包括配置数据和历史运行数据。 $\Delta L_i(t)$ 为在时隙*t*DTs同步需要更新的数据。 $S_i(t)$ 为设备运行状态,例如温度、轨迹、姿态等。 \mathcal{V}_i 是行为模型集合,模拟设备可能出现的各种行为。网络切片 $j \in \mathcal{J}$ 服务的多个DTs组成的网络表示为

$$DT^j = \left\{ \sum_{i \in \mathcal{I}_j} DT^d(i), CI_{NS}^j, SM_{NS}^j \right\} \quad (1)$$

其中 $\sum_{i \in \mathcal{I}_j} DT^d(i)$ 为划分到切片 $j \in \mathcal{J}$ 的DTs集合, CI_{NS}^j 为数字孪生网络(Digital Twin Network, DTN)信道状态, SM_{NS}^j 为切片 $j \in \mathcal{J}$ 中实例化DTs需要的软件模块。

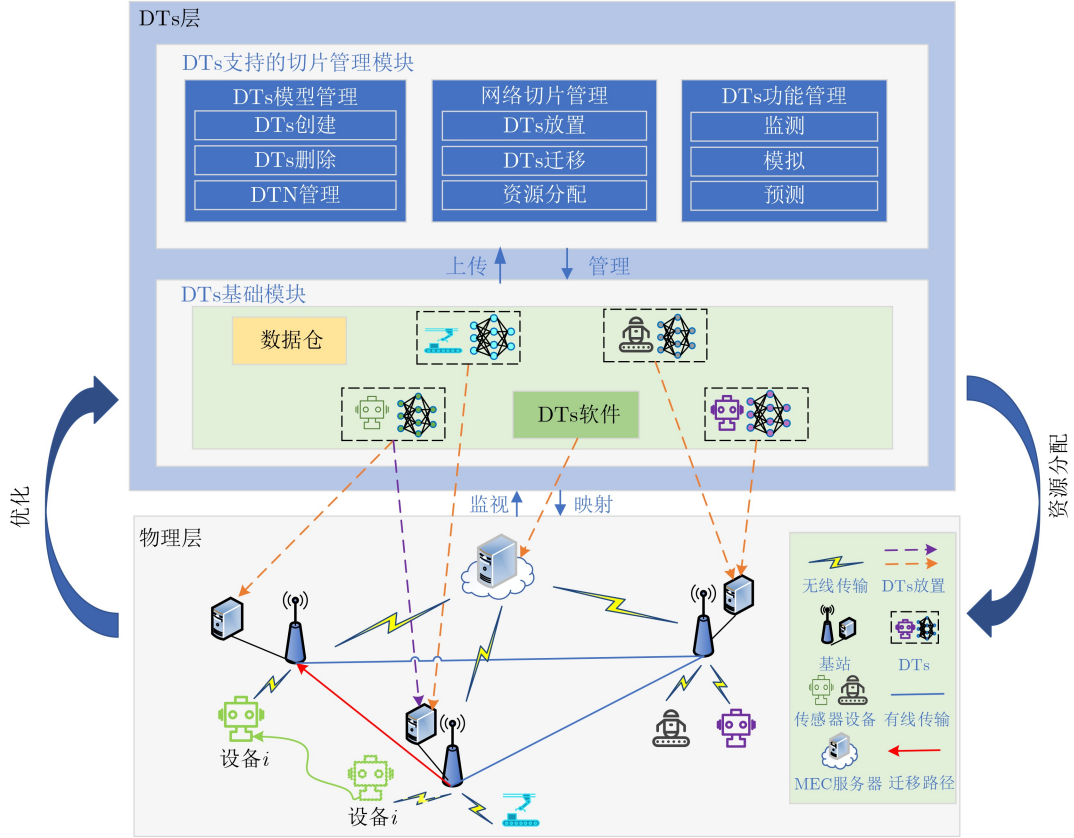


图 1 DTs支持的RAN切片场景

2.3 双时间尺度模型

为了避免频繁的切片重配置与DTs迁移导致成本过高及资源浪费，同时也更加符合现实世界的业务场景，本文在双时间尺度框架下对系统进行建模。假设边缘处存在 M 个基站，其集合表示为 $\mathcal{M} = \{1, \dots, m, \dots, M\}$ 。在基站的覆盖范围内，存在 J 个切片为DTs提供隔离，其集合表示为 $\mathcal{J} = \{1, \dots, j, \dots, J\}$ 。存在 I 个需要建立DT的用户设备，其集合表示为 $\mathcal{I} = \{1, \dots, i, \dots, I\}$ 。小时间尺度定义为小时间间隔 (Small Time Interval, STI)，用指标 $t \in \{1, 2, \dots\}$ 表示，STI的持续时间为 τ ；大时间尺度定义为大时间间隔 (Large Time Interval, LTI)，用指标 $k \in \{1, 2, \dots\}$ 表示，LTI的持续时间为 ΔT 个STI。在每个LTI中，系统将重新调整切片预留的无线资源并完成DTs的放置与迁移；在每个STI中，为完成DTs的传感数据同步，系统会进行细粒度资源分配。

2.3.1 小时间尺度模型

由 I 个传感设备、 M 个基站和1个云服务器组成的边缘网络系统中，每个传感设备 i 测量一个对应的线性时不变(LTI)过程 $\mathbf{x}_i(t)$ 的输出。所考虑的工业环境的状态基于如式(2)的离散时间动力学，随时间演变

$$\left. \begin{aligned} \mathbf{x}_i(t+1) &= \mathbf{A}\mathbf{x}_i(t) + \mathbf{w}_i(t) \\ \mathbf{y}_i(t) &= \mathbf{C}_i\mathbf{x}_i(t) + \mathbf{v}_i(t) \end{aligned} \right\} \quad (2)$$

其中， $\mathbf{x}_i(t)$ 是由传感设备 i 在时隙 t 获得的系统状态， $\mathbf{y}_i(t)$ 是传感设备 i 在时隙 t 的测量值；并且 \mathbf{A} 表示系统矩阵， \mathbf{C}_i 是传感设备 i 的测量矩阵；状态干扰噪声 $\mathbf{w}_i(t)$ 和测量噪声 $\mathbf{v}_i(t)$ 是均值为零并且方差为 $Q_{\mathbf{w}_i}$ 和 $Q_{\mathbf{v}_i}$ 的高斯白噪声。

本文考虑子载波间干扰，定义设备传输传感数据过程中的信干噪比 (Signal to Interference plus Noise Ratio, SINR) 为

$$\text{SINR}_i(t) \triangleq \frac{p_i(t)h_i(t)d_i^{-\vartheta}(t)}{I_- + \sigma^2} \quad (3)$$

其中， σ^2 为噪声功率， I_- 为其他设备的干扰，设备的传输功率和信道状态分别表示为 $p_i(t)$ 与 $h_i(t)$ ， $d_i^{-\vartheta}(t)$ 为路径损耗， $d_i(t) = \sqrt{(x_i^{\text{loc}}(t) - x_m)^2 + (y_i^{\text{loc}}(t) - y_m)^2}$ 表示设备 i 到基站 m 的距离，其中， $(x_i^{\text{loc}}(t), y_i^{\text{loc}}(t))$ 是传感设备 i 在时隙 t 的坐标位置， (x_m, y_m) 是基站 m 的坐标位置。

因此，设备 i 的上行链路传输速率为： $R_i(t) = b_i(t) \log_2(1 + \text{SINR}_i(t))$ ，其中 $b_i(t)$ 为分配给设备 i 的带宽。此外，当设备向DTs同步传感信息时，对延迟有着严格的约束，故假设当设备传输速率 $R_i(t)$

不小于阈值 R_j ，则切片 j 提供的通信资源满足设备QoS要求，表示为

$$R_i(t) \geq R_j, \forall i \in \mathcal{I}_j, \forall j \in \mathcal{J} \quad (4)$$

传输的可靠性与时延会直接影响估计性能，因此本文引入信息年龄(Age of Information, AoI)表达传感数据在到达基站之前由于不可靠的无线传输经历的估计延迟。首先，用二进制变量 $q_i(t)$ 来表示是否在时隙 t 传感设备 i 发送的信息能经受住干扰，即当接收到的SINR大于设定阈值 Γ^{th} 时，数据能够成功在基站侧解码，则 $q_i(t) = 1$ ，反之亦然， $q_i(t) = 0$ ，有

$$q_i(t) = \begin{cases} 1, & \text{SINR}_i(t) > \Gamma^{\text{th}} \\ 0, & \text{其它} \end{cases} \quad (5)$$

其次，考虑到传输速率对包成功到达基站的影响，用二进制变量 $\theta_i(t)$ 来表示设备 i 数据包 $l_i(t)$ 能否在时隙 t 内成功传输，当 $\theta_i(t) = 1$ 时，表示传感数据能在时刻内成功发送，反之亦然， $\theta_i(t) = 0$ ，有

$$\theta_i(t) = \begin{cases} 1, & R_i(t)\tau \geq l_i(t) \\ 0, & \text{其它} \end{cases} \quad (6)$$

在时隙 t ，使用AoI来衡量设备 i 产生的数据包经历的延迟，公式为

$$\tau_{i,\text{AoI}}(t) = \begin{cases} d_i(t)/\tau, & \theta_i(t) = 1 \\ \tau_{i,\text{AoI}}(t-1) + 1, & \text{其它} \end{cases} \quad (7)$$

其中， $d_i(t) = l_i(t)/R_i(t)$ 表示在时隙 t 内数据包完成传输所经历的延迟。

系统使用卡尔曼滤波器进行状态估计，状态估计被所经历的AoI延迟以及2元变量 $\theta_i(t)$ 以及 $q_i(t)$ 高度影响，公式为

$$\hat{\mathbf{x}}_i(t) = \begin{cases} \mathbf{A}\hat{\mathbf{x}}_i(t-1) + \mathbf{F}_i(t)(\mathbf{y}_i(t - \tau_{i,\text{AoI}}(t)) \\ \quad - \mathbf{C}_i\mathbf{A}\hat{\mathbf{x}}_i(t-1)), & q_i(t)\theta_i(t) = 1 \\ \mathbf{A}\hat{\mathbf{x}}_i(t-1), & \text{其它} \end{cases} \quad (8)$$

其中， $\hat{\mathbf{x}}_i(t)$ 是由设备 i 传感数据在基站估计的状态， $\mathbf{F}_i(t)$ 是卡尔曼增益。如果 $q_i(t)\theta_i(t) = 1$ ，则传感信息在时隙 t 内到达并成功解码，通过标准卡尔曼滤波器在基站更新状态。否则通过之前的估计来获得状态更新。

估计误差 $\mathbf{e}_i(t) = \mathbf{x}_i(t) - \hat{\mathbf{x}}_i(t)$ 能够评估感知信息状态估计的性能，假设基站处接收到最新的状态信息为 $\mathbf{x}_i(t - \tau_i(t))$ ，则式(2)重新表示为 $\mathbf{x}_i(t) = \mathbf{A}^{\tau_i(t)}\mathbf{x}_i(t - \tau_i(t)) + \sum_{o=1}^{\tau_i(t)} \mathbf{A}^{o-1}\mathbf{W}_i(t-o)$ 。为了让DTs精确的反映实体的状态和性能，本文定义 M_i^* 为DTs同步能容忍的最大估计MSE，因此在DTs同步过程中状态估计误差 χ_i 必须满足如式(9)的条件

$$\chi_i = \mathbb{E} \left\{ \|\mathbf{e}_i(t)\|_2^2 \right\} \leq M_i^*, \forall i \in \mathcal{I} \quad (9)$$

2.3.2 大时间尺度模型

首先，本文考虑切片成本。在DTs同步过程中，切片为其预留通信资源，则切片的成本来自于每个LTI通信资源调整比例的支出，因此，在第 k 个LTI中切片 j 成本为

$$F_{j,\text{op}}^k = c_{\text{sl}} \cdot |\alpha_{j,k} - \alpha_{j,k-1}| \quad (10)$$

其中， $F_{j,\text{op}}^k$ 为切片在第 $k-1$ 个LTI到第 k 个LTI带宽资源变化的成本函数， c_{sl} 为片缩放的单位成本， $\alpha_{j,k}$ 与 $\alpha_{j,k-1}$ 分别表示在第 k 个和第 $k-1$ 个LTI中切片 j 占用的频带资源百分比。

其次，本文考虑DTs迁移及实例化成本。假设设备处于多个基站的覆盖范围内，为了确保DTs能够实时监测设备并实现信息的同步，系统会自适应地将数字孪生迁移到新的基站，迁移成本可以表示为

$$C_{j,\text{mig}}^k = c_{\text{mig}} \cdot F_{j,\text{mig}}^k \quad (11)$$

其中， c_{mig} 为DT迁移的单位成本， $F_{j,\text{mig}}^k = \sum_{i \in \mathcal{I}_j} \sum_{m \in \mathcal{M}} \sum_{m' \in \mathcal{M}} 1\{\gamma_{i,m}^k \neq \gamma_{i,m'}^{k-1}\} \cdot L_i \cdot d(m', m)$ 为两个连续时隙的DTs的迁移函数，二进制变量 $\gamma_{i,m}^k$ 表示在第 k 个LTI中是否让设备 i 关联到基站 m 同步DTs数据， L_i 为孪生体迁移的数据包大小， $d(m', m)$ 为基站 m' 到基站 m 有线传输的距离。

当DTs完成迁移后，需要从MEC服务器下载对应软件模块支持实例化新的DTs，实例化成本为

$$C_{j,\text{ins}}^k = c_{\text{ins}} \cdot F_{j,\text{ins}}^k \quad (12)$$

其中， c_{ins} 为实例化DT的单位成本， $F_{j,\text{ins}}^k = \sum_{i \in \mathcal{I}_j} \sum_{m \in \mathcal{M}} \sum_{m' \in \mathcal{M}} 1\{\gamma_{i,m}^k \neq \gamma_{i,m'}^{k-1}\} \cdot D_{\text{soft}} \cdot d(m, \text{MEC})$ 为在两个连续LTI中DTs从基站 m' 到迁移到基站 m 后的部署成本函数， D_{soft} 为实例化需要软件模块的大小， $d(m, \text{MEC})$ 为基站 m 到MEC服务器的距离。

2.4 双时间尺度优化问题表述

本文提出一种基于智能分层切片技术的DTs同步策略，该策略不但能够掌握流量的长期变化趋势，且能够捕捉实时的任务动态。根据时间尺度的不同，同步策略 π 被分为上层控制策略 π_U 和下层控制策略 π_D 。上层控制器在第 k 个LTI根据前一LTI的DTs资源需求情况以及下层策略执行同步任务的反馈，实施策略 π_U 来为各个切片预留频带资源以及调整DTs的放置与迁移。下层控制器在上层控制策略 π_U 的约束下，在小时间尺度 t 调整功率、带宽，降低估计失真，提高同步任务对动态系统的鲁棒性和资源限制的适应性。

2.4.1 上层控制策略

上层控制器通过观察前一个LTI服务的变化情况，在下一个LTI预留切片需要的频谱资源以及完成DTs的放置。 $S_k = \{S_{j,k} \mid \forall j \in \mathcal{J}\} \in \mathcal{S}$ 表示在第 k 个LTI的状态，其中， \mathcal{S} 为所有可能出现的上层状态的有限集合； $S_{j,k} = \{\alpha_{j,k-1}, \gamma_{j,k-1}, \bar{V}_{j,k-1}, U_{j,k}, C_{j,k-1} \mid \forall j \in \mathcal{J}\}$ 为切片 j 的上层状态， $\alpha_{j,k-1}$ 是在第 $k-1$ 个LTI内切片 j 的频谱划分比例， $\gamma_{j,k-1}$ 是在第 $k-1$ 个LTI内切片 j 中所有DTs的关联集合， $\bar{V}_{j,k-1}$ 是在第 $k-1$ 个LTI内切片 j 中平均任务成功到达率， $U_{j,k}$ 是第 k 个LTI内切片 j 中的任务数， $C_{j,k-1}$ 是切片 j 在第 $k-1$ 个LTI内的效用。因此，上层控制策略 π_U 为网络的全局状态到频谱资源比例划分以及设备与基站的关联(DTs放置)的映射： $\pi_U: S_k \rightarrow C_k$ ，其中， $C_k = \{\alpha_{j,k}, \gamma_{j,k} \mid \forall j \in \mathcal{J}\} \in \mathcal{C}$ 为切片上层资源配置， \mathcal{C} 是所有上层资源配置方案的集合。因此上层控制策略能让每个切片中设备共享资源，实现DTs灵活部署，提高片的资源利用效率。

2.4.2 下层控制策略

下层状态信息表示为 $X_t = \{X_{j,t} \mid \forall j \in \mathcal{J}\} \in \mathcal{X}$ ，其中， \mathcal{X} 为所有可能出现的上层状态的有限集合，下层片 j 状态信息表示为 $X_{j,t} = \{\bar{\tau}_{j,t-1}, \chi_{i,t-1} \mid i \in \mathcal{I}_j\}$ ，其中， $\bar{\tau}_{j,t-1}$ 和 $\chi_{i,t-1}$ 分别为观测到上一时隙的片 j 中任务平均AoI和估计误差。因此，在大时间尺度的上层控制策略 π_U 被实施的情况下，小时间尺度的下层控制策略 $\pi_D: (X_t, C_k) \rightarrow E_t, k\Delta T \leq t < (k+1)\Delta T$ 是将观测到上一时隙的片 j 的状态信息的映射，其中，小时间尺度的资源分配方案为 $E_t = \{E_{j,t} \mid \forall j \in \mathcal{J}\} \in \mathcal{E}$ ， \mathcal{E} 为所有可能出现的下层方案的有限集合，每个切片的资源分配方案为 $E_{j,t} = \{P_{j,t}, B_{j,t} \mid \forall j \in \mathcal{J}\} \in \mathcal{E}$ ，其中， $P_{j,t} = \{p_i(t) \mid \forall i \in \mathcal{I}_j\}$ 与 $B_{j,t} = \{b_i(t) \mid \forall i \in \mathcal{I}_j\}$ 为小时间尺度中切片 j 配置的功率及带宽。因此，在STI中策略 π_D 进行细粒度的资源配置来降低DTs同步数据的估计误差。

2.4.3 问题建模

为了最大化系统收益的同时控制系统成本，首先定义切片 j 在第 k 个LTI的总收入函数 C_j^k 为下层的收益函数 $C_{j,es}^k(X_{j,t}, E_t)$ 和上层DTs总成本函数 $C_{j,ad}^k(S_{j,k})$ 的差值： $C_j^k = C_{j,es}^k(X_{j,t}, E_t) - C_{j,ad}^k(S_{j,k})$ ，其中，切片 j 的上层DTs总成本函数 $C_{j,ad}^k(S_{j,k})$ 由片缩放成本 $C_{j,op}^k$ 及DTs同步成本 $C_{j,dig}^k$ 组成，即 $C_{j,ad}^k(S_{j,k}) = \beta_{j,1}C_{j,op}^k + \beta_{j,2}C_{j,dig}^k$ 。由于系统获得收益与用户传感信息满意度呈正相关，故在第 k 个LTI系统收益为

$$C_{j,es}^k = \sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{I}_j} \sum_{t=k\Delta T+1}^{(k+1)\Delta T} \beta_{j,3} c_{es} \gamma_{i,m}^k (M_i^* - \chi_i(t)),$$

c_{es} 为单位收益， $\beta_{j,1}$ 、 $\beta_{j,2}$ 和 $\beta_{j,3}$ 为权重系数。

本文旨在满足无线资源调度约束的前提下，寻求最佳的智能分层切片同步策略的最优性能，使最大化系统收益的同时控制系统成本，则优化问题可以表述为

$$\mathcal{P}: \min_{\pi = \{\pi_D, \pi_U\}} \left\{ \frac{1}{T} \sum_{j=1}^J \sum_{k=0}^T C_j^k \right\},$$

$$\text{s.t.} \begin{cases} \text{式(4), 式(9)} \\ \text{C1: } (\alpha_{j,k}, \alpha_{-j,k}) \in [0, 1], k = 1, 2, \dots \\ \text{C2: } \alpha_{j,k} + \alpha_{-j,k} \in [0, 1], k = 1, 2, \dots \\ \text{C3: } \gamma_{i,m}^k \in \{0, 1\}, \forall m \in \mathcal{M}, \forall i \in \mathcal{I}_j, \\ \quad k = 1, 2, \dots \\ \text{C4: } \sum_{m \in \mathcal{M}} \gamma_{i,m}^k \leq 1, i \in \mathcal{I}_j, \forall j \in \mathcal{J}, \\ \quad k = 1, 2, \dots \\ \text{C5: } \Pr \{R_i(t) \geq l_i(t)/\tau\} \geq 1 - \frac{1}{\rho^2(A)}, \\ \quad \forall i \in \mathcal{I}_j, \forall j \in \mathcal{J} \\ \text{C6: } \sum_{i \in \mathcal{I}_j} b_i(t) \leq \alpha_j^k B, \forall j \in \mathcal{J} \\ \text{C7: } \sum_{m \in \mathcal{M}} \gamma_{i,m}^k p_i(t) \leq P_{\max}, \forall i \in \mathcal{I}_j, \forall j \in \mathcal{J} \end{cases} \quad (13)$$

C1表示切片 j 占用的频带资源百分比和其余切片占用频带资源百分比应在0~1之间；C2表示切片 j 占用资源与其余切片占用资源的比率总和限制在0~1之间；C3代表DTs关联变量为2元变量；C4表示确保每个传感设备最多与1个基站建立关联；C5表示状态估计收敛的约束；C6表示设备带宽资源之和不能高于切片的带宽资源；C7表示发射功率的约束。

3 基于DDQN-PER-MADDPG的分层解决方案

由于式(19)属于混合整数非线性问题，加之双时间尺度问题的耦合性，使用传统方法难以求解。因此，本文把问题拆分为两个子问题，将其转化为双马尔可夫决策过程(Markov Decision Process, MDP)，上层控制过程转化为无限长度MDP，下层控制过程转化有限长度MDP，然后利用分层强化学习求解。针对小时间尺度问题，使用PER-MADDPG算法学习下层控制策略求解；在收敛的下层策略基础上，针对大时间尺度问题，使用DDQN学习上层控制策略求解。

3.1 双时间尺度MDP

MDP可由 $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P})$ 组成的4元组来描述, 其中, \mathcal{S} 是状态空间, \mathcal{A} 是动作空间, \mathcal{P} 是状态转移概率, \mathcal{R} 是奖励。在MDP中, 智能体忽略了历史状态信息, 仅考虑当前状态来选择动作, 并在与环境完成交互后获得回报, 每个回合的状态、行动和回报的顺序组成了迹 ζ 。在DRL中, 智能体的目标是找到一种最优策略 $\pi^* = \arg \max_{\pi} \mathbb{E}[\mathcal{R} | \pi]$, 其中策略 π 是在给定状态下执行动作的概率分布, $R = \sum_{t=0}^{T-1} \gamma^t r(t+1)$ 是在有限长度的情况下MDP的累计回报; 动作价值函数 $Q_{\pi}(s, a) = \mathbb{E}_{\pi}[\mathcal{R} | s(t) = s, a(t) = a]$ 是对智能体在当前状态下采取某个动作后可获得的潜在收益的评估, 如果智能体的策略能够得到最优的动作价值函数 $Q_{\pi}^*(s, a)$, 那么该策略是最优策略, 这是是DRL求解的最重要步骤; 因此, 上下两层的动作价值函数如下分析。

在上层给出了资源配置方案 $C_k \in \mathcal{C}$ 的基础上, 从 t 开始到第 k 个LTI结束, 下层动作价值函数定义为智能体根据观察到的状态 $s(t)$, 学习策略 π_D 来执行动作 $a(t)$ 获得的期望累计奖励

$$Q_L(X_t, E_t | C_k) = \mathbb{E}_{\pi_D} \left[\sum_{t'=t}^{(k+1)\Delta T-1} \sum_{\forall j \in \mathcal{J}} C_{j,t'}^{\text{es}}(X_{j,t'}, E_{t'}) | X_t, E_t, C_k \right] \quad (14)$$

在上层资源配置方案 $C_k \in \mathcal{C}$ 下, 智能体根据策略 π_D 完成与环境的交互, 获得期望累计奖励 $Q_L(X_t, E_t | C_k)$, 因此智能体在第 k 个LTI获得的期望累积回报为 $J_L(\pi_D | C_k) = \mathbb{E}_{\zeta} [Q_L(X_{k\Delta T}, E_{k\Delta T} | C_k)]$, 其中, $\zeta = (X_{k\Delta T}, E_{k\Delta T}, \dots, X_{(k+1)\Delta T}, E_{(k+1)\Delta T})$ 为在第 k 个LTI中所有智能体与环境交互得到的迹。

在第 k 个LTI开始时, 系统首先执行上层策略 π_U , 因此上层动作价值函数为

$$Q_U(S_k, C_k) = \mathbb{E}_{\pi_U} \left[\left(\sum_{k'=k}^{\infty} \lambda^{k'-k} (J_L(\pi_D | C_{k'})) - \sum_{i \in \mathcal{I}} C_{i,k'}^{\text{ad}}(S_{i,k'}) \right) | S_k, C_k, \pi_D \right] \quad (15)$$

因此, 目标函数 \mathcal{P} 重新定义为 $J(\pi) = \mathbb{E}[Q_U(S_k, C_k) | \pi_U, \pi_D] = \mathbb{E}_{\pi_U} \left[\sum_{k=0}^{\infty} \lambda^k \left(J_L(\pi_D | C_k) - \sum_{i \in \mathcal{I}} C_{i,k}^{\text{ad}}(S_{i,k}) \right) \right]$

3.2 子目标问题表述

在上层资源配置方案 $C_k \in \mathcal{C}$ 的约束下, 下层的目标是学习得到最大期望奖励的最优下层策略

π_D^* , 目标子问题 $\mathcal{P}_1: \pi_D^* = \arg \max_{\pi_D} \{J_L(\pi_D | C_k) | C_k \in \mathcal{C}\}$, 受式(4)、式(9)、C6与C7的约束。基于收敛的下层策略 π_D^* , 求解目标子问题 \mathcal{P}_2 来获得最优的上层策略 $\pi_U^* = \arg \max_{\pi_U} \{J(\pi) | \pi_D^*\}$, 受C1~C5的约束。

3.3 基于PER-MADDPG的下层控制算法

MADDPG算法能够有效处理整个网络系统的竞争和合作关系, 采用策略集合的思想, 使得智能体能够适应彼此的策略。每个基站被视为一个智能体, 具体的MDP元组定义为智能体 m 在当前时隙 t 观测的状态 $s_m(t) = \{X_m(t)\}$, 其中, $X_m(t) = \{\bar{\tau}_{j,t}, \chi_{m,i,t} | i \in \mathcal{I}_j, \forall j \in \mathcal{J}\}$; 智能体在MEC服务器中运行的MADDPG算法采用“集中式训练, 分布式执行”的评论家-行动家结构。算法中每个智能体根据自身观测的信息, 独立采取动作, 即 $a_m = \pi_m^D(s_m)$ 。评论家根据全局信息评估行动家的行为, 并且目标网络的参数定期地进行软更新。其中 $\pi_D = \{\pi_1^D, \dots, \pi_m^D, \dots, \pi_M^D\}$ 为 M 个智能体的策略, 并且 $\theta^Q = \{\theta_1^Q, \dots, \theta_m^Q, \dots, \theta_M^Q\}$ 为评论家网络参数, $\theta^E = \{\theta_1^E, \dots, \theta_m^E, \dots, \theta_M^E\}$ 为行动家网络参数。

评论家网络的损失函数为样本真实 Q 值和估计 Q 值的均方误差, 即

$$\mathcal{L}_m(\theta_m^Q) = \mathbb{E} \left[(Q_m(s_m, a_m | C_k, \theta_m^Q) - y_m)^2 \right] \quad (16)$$

其中, 样本真实 Q 值 $y_m = r_m + \gamma Q'_m(s'_m, a'_m | C_k, \theta_m^Q)$, s'_m 和 a'_m 分别为 s_m, a_m 下一状态和动作。

智能体 m 朝着最大化期望奖励 $J_m(\pi_m^D | C_k) = \mathbb{E}_s [Q_m(s_m, a_m | C_k)]$ 采取对应动作, 根据策略梯度, 行动家网络进行更新, 有

$$\nabla_{\theta_m^E} J \approx \frac{1}{Z} \sum_z \nabla_{\theta_m^E} \pi_m^D(s_m^z) \nabla_{a_m} Q_m(s_m^z, a_m^z) \Big|_{a_m = \pi_m^D(s_m^z)} \quad (17)$$

其中, Z 代表参与一次本地网络训练需要的经验的数量大小, z 为经验的指数。

评论家网络参数 θ_m^Q 和行动家网络参数 θ_m^E 分别以最小化的损失函数和计算策略梯度来获得更新, 目标网络参数采用软更新方式获得更新。

引入PER机制后, 智能体会频繁采样对网络更新更重要的经验, 从而提高网络学习效率。时间差分误差(Temporal Difference error, TD-error)的绝对值可作为评定优先级的依据。经验 w 的TD-error表示为 $\delta_w = r + \gamma Q'_r(s', a') - Q_r(s, a)$ 。TD-Error绝对值差值越大, 期望动作值的正相关性越强, 故

应增加对这些经验采样的频率，设置更高的优先级。采样概率 $P(w)$ 定义为 $P(w) = D_w^\nu / \sum_z D_z^\nu$ ，其中， $D_w = 1/\text{rank}(w) > 0$ ， $\text{rank}(w)$ 为基于TD-error绝对值 $|\delta_w|$ 的经验先后顺序， ν 为优先级的使用程度，通常控制在0-1之间，其中 $\alpha = 0$ 代表均匀采样。TD-error值大的经验因其具有更高的概率，则会被更频繁的采样，然而，这种情况导致批数据的分布与回放池中经验分布不同，不具有代表性，甚至可能造成训练过程网络剧烈震荡，与预期的收敛情况不同。因此在训练过程中权重设置为 $\omega_w = 1/S^\beta \cdot P(w)^\beta$ ，其中 S 为经验回放池的大小， $P(w)$ 为采样概率， β 为用于调整采样经验权重的参

$$\mathcal{L}(\varphi) = \mathbb{E}_s \left[\left(Q(s(k), a(k), \mu) - \left(r + \gamma_U Q \left(s(k+1), \arg \max_{a(k)} Q(s(k+1), a(k), \mu), \mu_- \right) \right) \right)^2 \right) \quad (18)$$

其中， μ_- 代表目标网络的参数。Q网络使损失函数梯度最小化来更新参数，损失函数的梯度计算公式为

$$\nabla_\mu \mathcal{L}(\mu) = \mathbb{E} \left[r(k) + \gamma_U Q \left(s(k+1), \arg \max_{a(k)} Q(s(k+1), a(k), \mu), \mu_- \right) - Q(s(k), a(k), \mu) \right] \nabla_\mu Q(s(k), a(k), \mu) \quad (19)$$

智能体采用 ε -贪婪策略 $\pi_U(\mathbf{C}_k | \mathbf{S}_k)$ 有一定概率随机选择动作进行探索，使其避免陷入局部最优解；智能体有 $1 - \varepsilon$ 的概率更着重眼前的利益，选择最大Q值对应的动作；智能体有 $\varepsilon/|\mathbf{C}| - 1$ 的概率

数。基于PER-MADDPG的下层控制算法如算法1所示。

3.4 基于DDQN的上层控制算法

上层控制器根据网络状态的变化情况，配置上层资源，让系统收益最大。在表示MDP的元组中，算法的观测状态与上层状态相同： $s(k) = S_k$ ；动作集为 $a(k) = \mathbf{C}_k$ ；奖励为 $r(k) = \sum_{j \in \mathcal{J}} C_j^k$ 。

DDQN利用目标网络与训练网络形成的双网结构，在解决DQN估计奖励高于实际奖励的问题同时，也让网络结构更稳定。双网结构的损失函数能衡量策略好坏，让智能体朝着最优策略方向训练，表示为

更关心长远利益，随机选择动作进行探索。基于DDQN的上层控制算法如算法2所示。

4 仿真结果与分析

首先对系统仿真参数进行说明。仿真软件环境基于Python和Tensorflow。在 $[0,10] \text{ km} \times [0,10] \text{ km}$ 的工业物联网网络拓扑下，基站 $M=4$ ，用户设备数 $I=20$ 。根据用户设备的业务需求将切片划分为切片1和切片2，其中划分到切片1的用户设备对DTs同步的速率有更高的需求。其次是算法仿真参数设置说明。上层和下层控制算法参数均为独立配

算法1 基于PER-MADDPG的下层控制算法

输入：学习率 λ ，小批量大小 Z ，经验池 D_L ，参数 ν ，参数 β
输出：下层控制策略

- (1) for episode = 1 ~ E_L do
- (2) 所有代理都观察初始环境状态 s
- (3) for step = 1 ~ T_L do
- (4) 所有智能体按照策略采取行动 \mathbf{a} 并添加环境噪声 N_t
- (5) 与环境交互获得各自惩罚奖励 r 以及跳转到下一状态 s' ，并把经验 (s, a, r, s') 存储在 D_L
- (6) for 智能体 $m = 1 \sim M$ do
- (7) for $z = 1 \sim Z$ do
- (8) 从经验池 D_L 中以 $P(k)$ 的概率抽取样本 w
- (9) 根据实际奖励计算TD-error δ_w 以及计算权重 ω_w
- (10) 根据绝对TD-error $|\delta_w|$ 更新样本 w 基于排名的优先级
- (11) end for
- (12) 计算全局 $\mathcal{L}(\theta_m^Q) = \frac{1}{Z} \sum_z \omega_w \delta_w^2$ ，并最小化 $\mathcal{L}(\theta_m^Q)$ 来更新评论家网络
- (13) 计算策略梯度 $\nabla_{\theta_m^E} J$ ，更新行动家网络
- (14) end for
- (15) 更新智能体的目标网络
- (16) end for
- (17) end for

算法2 基于DDQN的上层控制算法

输入：概率分布 ψ ，探索概率 ε ，小批量大小 B ，采样数据的学习回合数
输出：上层控制策略

- (1) 初始化神经网络参数
- (2) for episode = 1 ~ E_U do
- (3) 观察环境获得初始观测值 s
- (4) for step=1 ~ T_U do
- (5) 根据 ε -贪婪策略选择动作 a ，即选择探索动作还是最大Q值对应动作
- (6) 控制器与环境交互获得 r 并跳转到下一状态 s' ，并采集经验 (s, a, r, s') 放到回放池 D_U
- (7) 从回放池 D_U 抽取一批经验
- (8) 计算梯度 $\nabla_\mu \mathcal{L}(\mu)$ ，完成网络参数 μ 反向更新
- (9) 每隔 G 步，复制网络参数 μ 给目标网络参数 μ_-
- (10) end for
- (11) end for

置。对于下层控制PER-MADDPG算法,评论家网络和行动家网络均设置3个隐藏层且每个隐藏层均含有128个神经,并采用Relu作为激活函数。对于上层控制DDQN算法,网络设置两个全连接层,每一层含有64个神经元,激活函数为Relu。具体仿真参数总结于表1。

(1)算法性能分析:图2将本文方案与其他3种算法进行比较,对比算法设置为双深度Q网络-多智体深度确定性梯度(Double Deep Q Network - Multi-Agent Deep Deterministic Policy Gradient, DDQN-MADDPG)算法、双深度Q网络-深度确定性梯度(Double Deep Q Network - Deep Deterministic Policy Gradient, DDQN-DDPG)算法以及非分层(Non-Hierarchical, Non-H)算法。Non-H算法采用单一时间尺度的框架,并且取消了切片的划分,利用DDPG算法对资源集中配置。本文方案与DDQN-MADDPG算法总奖励是接近的,但本文方案采用PER机制,具有更快的收敛速度和更高的学习效率,在1 300轮训练后收敛,后者在2 000轮训练后收敛。DDQN-DDPG算法采用单一智能体与环境交互,不考虑多智体之间的合作,因此具有更快收敛速度,在950轮训练后收敛,然而奖励是低于多智体算法的。Non-H算法难以处理混合和离散的高维动作空间,且DDPG仅支持连续动作空间,因此性能是最差的。

表1 仿真参数设置

参数	值	参数	值
基站数量	4	下层评论家/ 行动家学习率	0.01/0.001
IoT设备	20	上层/下层折扣因子	0.9/0.95
带宽	1.8 MHz	上层/下层最小批	512/32
每个LTI的长度(τ)	100 ms	单位DT迁移/实例化成本	15/15
每个STL的长度(ΔT)	5 s	切片1/切片2速率阈值	600/300
最大传输功率	40 mW	上层贪婪率	0.1

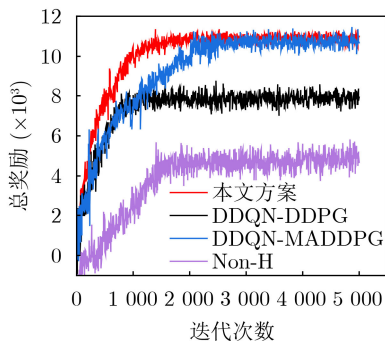


图2 不同算法奖励对比

(2)资源分配策略:图3展示在15步内的DTs同步成本的变化。其中Random算法是上层采用LSTM方法对带宽资源进行预测,下层采用DDPG算法来进行资源分配,当智能体观测到环境变化后就将DTs随机地迁移到任意基站中;同理,Nearest算法是把DTs迁移到离传感设备最近的基站。其中Random算法的DTs同步成本是最高的,因为Random算法在设备的状态发生改变时总是随机地去迁移原来的对应的DTs,因此造成不必要额外成本。本文方案同步成本是高于其他两种算法的,因为分层DRL算法会协同决策,利用较高同步成本能够换取低AoI,保证同步的精确性。

AoI能表征数据的可靠性与时延对DTs的传感信息同步的影响。图4为不同切片方案的AoI的累积分布函数。切片1中98%的平均AoI是分布在0~40 ms,切片2中几乎全部平均AoI低于80 ms。这表明不同切片方案能够满足不同的业务需求。切片1和切片2的平均AoI分别如图5(a)与图5(b)所示。其中,本文方案的平均AoI均小于其他3种方案,证明智能分层切片对DTs传感信息同步重要性以及相比于其他算法的优越性。DDQN-DDPG算法的平均AoI是高于本文方案,因为多智体通过相互协作能降低问题的复杂性,做出更优决策。Nearest算法的AoI是高于分层强化学习方法的,随机算法的平均AoI是最大的,这是因为分层强化学习使用分

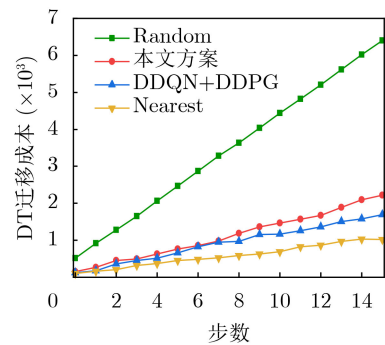


图3 不同算法DTs同步成本对比

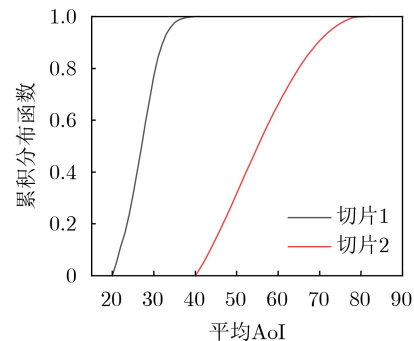


图4 不同切片方案的AoI累积分布函数对比

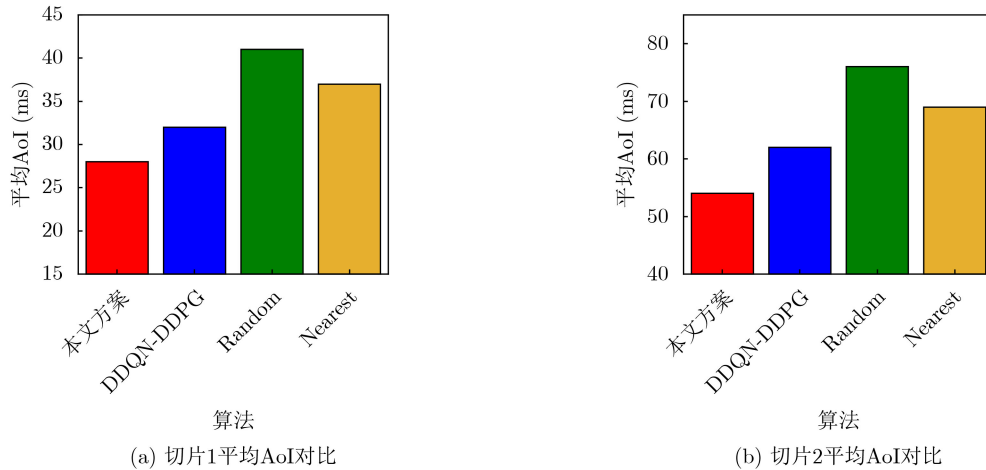


图 5 不同切片中不同算法AoI对比

层结构来学习不同层次的策略，得到更好的资源分配方案。

状态估计的MSE是衡量DTs同步信息的精确度指标，其越小代表估计的信息越精确。图6展示了不同算法的状态估计的MSE。如图6所示，本文方案状态估计MSE是最小的，因此本文方案能实现高精度的DTs传感信息同步；DDQN-DDPG算法MSE是大于本文方案，Random算法MSE是最高的。

5 结束语

本文提出了一种基于智能分层切片技术的数字孪生传感信息同步策略，首先设计了双时间尺度的系统模型，以最大化传感信息满意度和最小化切片重配置及DTs同步成本为目标，联合切片优化无线电资源配置以及DTs传感信息同步问题。其次为了求解双时间尺度的优化问题，本文将DRL求解过程转化为双MDP，利用DDQN-PER-MADDPG求解问题，并采用PER机制加快收敛。最后仿真结果表明所本文方案在收敛以及其他性能方面的优越性。

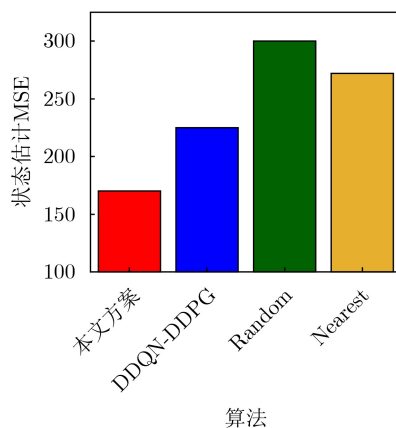


图 6 不同算法的MSE对比

参考文献

- [1] ZEB S, MAHMOOD A, HASSAN S A, *et al.* Industrial digital twins at the nexus of NextG wireless networks and computational intelligence: A survey[J]. *Journal of Network and Computer Applications*, 2022, 200: 103309. doi: 10.1016/j.jnca.2021.103309.
- [2] LIN Xingqin, KUNDU L, DICK C, *et al.* 6G digital twin networks: From theory to practice[J]. *IEEE Communications Magazine*, 2023, 61(11): 72–78. doi: 10.1109/MCOM.001.2200830.
- [3] KURUVATTI N P, HABIBI M A, PARTANI S, *et al.* Empowering 6G communication systems with digital twin technology: A comprehensive survey[J]. *IEEE Access*, 2022, 10: 112158–112186. doi: 10.1109/ACCESS.2022.3215493.
- [4] KHAN L U, SAAD W, NIYATO D, *et al.* Digital-twin-enabled 6G: Vision, architectural trends, and future directions[J]. *IEEE Communications Magazine*, 2022, 60(1): 74–80. doi: 10.1109/MCOM.001.21143.
- [5] WU Yiwen, ZHANG Ke, and ZHANG Yan. Digital twin networks: A survey[J]. *IEEE Internet of Things Journal*, 2021, 8(18): 13789–13804. doi: 10.1109/JIOT.2021.3079510.
- [6] LU Yunlong, HUANG Xiaohong, ZHANG Ke, *et al.* Low-latency federated learning and blockchain for edge association in digital twin empowered 6G networks[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(7): 5098–5107. doi: 10.1109/TII.2020.3017668.
- [7] LIU Tong, TANG Lun, WANG Weili, *et al.* Resource allocation in DT-assisted internet of vehicles via edge intelligent cooperation[J]. *IEEE Internet of Things Journal*, 2022, 9(18): 17608–17626. doi: 10.1109/JIOT.2022.3156100.
- [8] LU Yunlong, MAHARJAN S, and ZHANG Yan. Adaptive edge association for wireless digital twin networks in 6G[J]. *IEEE Internet of Things Journal*, 2021, 8(22): 16219–16230. doi: 10.1109/JIOT.2021.3098508.

- [9] SUI Tianju, YOU Keyou, and FU Minyue. Stability conditions for multi-sensor state estimation over a Lossy network[J]. *Automatica*, 2015, 53: 1–9. doi: [10.1016/j.automatica.2014.12.022](https://doi.org/10.1016/j.automatica.2014.12.022).
- [10] CHUKHNO O, CHUKHNO N, ARANITI G, *et al.* Placement of social digital twins at the edge for beyond 5G IoT networks[J]. *IEEE Internet of Things Journal*, 2022, 9(23): 23927–23940. doi: [10.1109/JIOT.2022.3190737](https://doi.org/10.1109/JIOT.2022.3190737).
- [11] LYU Ling, DAI Yanpeng, CHENG Nan, *et al.* AoI-aware co-design of cooperative transmission and state estimation for marine IoT systems[J]. *IEEE Internet of Things Journal*, 2021, 8(10): 7889–7901. doi: [10.1109/JIOT.2020.3041287](https://doi.org/10.1109/JIOT.2020.3041287).
- [12] WIJETHILAKA S and LIYANAGE M. Survey on network slicing for internet of things realization in 5G networks[J]. *IEEE Communications Surveys & Tutorials*, 2021, 23(2): 957–994. doi: [10.1109/COMST.2021.3067807](https://doi.org/10.1109/COMST.2021.3067807).
- [13] CHIANG Y, HSU C H, CHEN G H, *et al.* Deep Q-learning-based dynamic network slicing and task offloading in edge network[J]. *IEEE Transactions on Network and Service Management*, 2023, 20(1): 369–384. doi: [10.1109/TNSM.2022.3208776](https://doi.org/10.1109/TNSM.2022.3208776).
- [14] YE Feng, WANG Jie, LI Jiamin, *et al.* Intelligent hierarchical network slicing based on dynamic multi-connectivity in cell-free distributed massive MIMO systems[J]. *IEEE Transactions on Vehicular Technology*, 2023, 72(9): 11855–11870. doi: [10.1109/TVT.2023.3268822](https://doi.org/10.1109/TVT.2023.3268822).
- [15] KHAN L U, HAN Zhu, SAAD W, *et al.* Digital twin of wireless systems: Overview, taxonomy, challenges, and opportunities[J]. *IEEE Communications Surveys & Tutorials*, 2022, 24(4): 2230–2254. doi: [10.1109/COMST.2022.3198273](https://doi.org/10.1109/COMST.2022.3198273).
- 唐 伦: 男, 教授, 博士生导师, 研究方向为新一代无线通信网络、异构蜂窝网络、软件定义无线网络等。
- 李质萱: 女, 硕士生, 研究方向为数字孪生、网络切片、资源分配等。
- 文 雯: 女, 硕士生, 研究方向为虚拟化网络切片、资源分配、数字孪生等。
- 成章超: 男, 硕士生, 研究方向为车联网、数字孪生、深度强化学习等。
- 陈前斌: 男, 教授, 博士生导师, 研究方向为个人通信、多媒体信息处理与传输、下一代移动通信网络、异构蜂窝网络等。

责任编辑: 余 蓉