

doi: 10.16104/j.issn.1673-1891.2025.04.009

## 基于深度强化学习的语义通信动态资源调度优化

祖 婷, 伍 祥, 王国义

(安徽机电职业技术学院互联网与通信学院, 安徽 芜湖 241002)

**摘要:**为解决动态网络环境下多任务、多优先级流共存场景中,语义通信系统资源调度存在的效率低下与优先级保障不足的问题,本研究采用深度Q网络(deep Q-network, DQN)结合语义优先级和信道质量动态分配资源,提出基于深度强化学习的语义感知动态资源调度框架;通过引入经验回放和目标网络技术优化学习过程,提高模型稳定性与收敛效率。结果表明,所提方法在带宽利用率(96%)、传输延迟(8.98 ms)和高优先级流保障(关键帧准确率98.3%)上显著优于对比方案;语义恢复质量(2.73)虽略低于固定分配,但仍满足实时通信需求。所提方法通过语义优先级与DQN的动态优化机制,有效平衡了资源利用率与关键流服务质量,为复杂语义通信系统提供了高效资源调度方案。

**关键词:**语义通信;深度强化学习;动态资源调度;深度Q网络

中图分类号:TP391.1;TN929.5 文献标志码:A 文章编号:1673-1891(2025)04-0079-07

## Optimization of Dynamic Resource Scheduling for Semantic Communication Based on Deep Reinforcement Learning

ZU Ting, WU Xiang, WANG Guoyi

(School of Internet and Communication, Anhui Technical College of Mechanical and Electrical Engineering, Wuhu 241002, Anhui, China)

**Abstract:**To solve the problems of low efficiency and insufficient priority guarantee in semantic communication system resource scheduling in the scenario of multi task and multi priority flow coexistence in dynamic network environment, a deep Q-network (DQN) was used to dynamically allocate resources based on semantic priority and channel quality, and a semantic aware dynamic resource scheduling framework was proposed based on deep reinforcement learning. By introducing experience replay and target network technology to optimize the learning process, the stability and convergence efficiency of the model were enhanced. The results show that the proposed method significantly outperforms the comparison scheme in terms of bandwidth utilization (96%), transmission delay (8.98 ms), and high priority flow assurance (key frame accuracy of 98.3%). Although the semantic recovery quality (2.73) is slightly lower than fixed allocation, it still meets real-time communication requirements. The proposed method effectively balances resource utilization and critical flow service quality through the dynamic optimization mechanism of semantic priority and DQN, providing efficient resource scheduling solutions for complex semantic communication systems.

**Keywords:**semantic communication; deep reinforcement learning; dynamic resource scheduling; deep Q-network

收稿日期:2025-04-22

基金项目:基于深度学习的语义通信系统在动态数据环境下的应用研究(2024AH050212)。

第一作者简介:祖婷(1991—),女,安徽安庆人,讲师,硕士,主要研究方向为语义通信、人工智能。E-mail: 0120200014@ahcme.edu.cn。

## 0 引言

伴随信息通信技术的迅猛进步,传统的数据传输模式在处理动态繁杂的网络环境时,其局限性日益显现,特别是在物联网(internet of things, IoT)、5G 及未来 6G 网络这类场景里,大量非连续、动态的数据流对资源调度提出了更高要求<sup>[1]</sup>。语义通信作为一种突破传统通信范式的技术,通过直接传递信息的语义内涵,显著降低了数据冗余,提高了通信效率,为解决上述挑战提供了新思路<sup>[2]</sup>。近年来,强化学习(reinforcement learning, RL)被广泛应用于网络资源管理领域,其通过与环境的交互不断学习最优策略,展现了良好的动态适应能力。不过,传统的强化学习方法(如 Q 学习(Q-Learning)和 SARSA(State-Action-Reward-State-Action))在高维状态空间中的收敛速度较慢,难以解决实时且复杂的网络环境问题<sup>[3]</sup>。随着深度学习的发展,深度强化学习(deep reinforcement learning, DRL)通过结合深度神经网络和 RL 技术,能够在解决高维数据和复杂环境中具有较好的性能,已逐渐变成网络资源调度领域的研究重点<sup>[4-5]</sup>。

在语义通信跟网络资源调度领域,已经有大量的研究在努力提高通信效率和资源利用率。语义通信的关键理念最初由香农提出<sup>[6]</sup>,他着重指出信息传递时语义得准确、有效。近年来,由于人工智能技术的不断发展,语义通信受到了广泛的关注。张平等<sup>[7]</sup>通过基于深度学习的语义编码技术,提取并压缩信息的语义特征,减少数据传输量。其他研究对语义感知的动态适应性做了进一步研究,证明了语义优先级在资源优化中的关键意义<sup>[8]</sup>。传统的网络资源调度手段大多依照静态分配策略来进行,靠预定规则达成资源分配,但是这种方法在动态网络环境里存在较大的局限性<sup>[9]</sup>。随机分配与启发式方法(如遗传算法、模拟退火等)给资源优化带来了一些改进,但在复杂环境里,还是难以满足实时性

和效率的要求。近几年,基于模型优化的方法(如卷积神经网络等),给资源调度带来了新的方向<sup>[10]</sup>。

强化学习通过与环境的交互学习得到最优策略,在网络资源管理领域得到广泛应用。Ye 等<sup>[11]</sup>把 Q-Learning 方法应用在无线网络里的功率控制问题上,充分证明了它具有适应动态环境的能力。然而,传统强化学习在高维状态空间里的表现无法满足一些复杂任务的要求。Chen 等<sup>[12]</sup>通过借助深度 Q 网络(deep Q-network, DQN)优化异构网络中的资源分配问题,在提高带宽利用率的同时也增强了延迟控制能力。但是,语义通信系统的性能往往在很大程度上依赖于资源的合理调度,尤其是在多任务、多优先级流共存的场景中,所以如何动态优化资源分配来满足多样化需求仍然是一个亟待解决的问题。

综上,本文设计了一种通过结合语义优先级和信道质量的奖励机制,用于优化深度强化学习模型动态调整资源分配策略,提出了基于 DQN 的动态资源调度框架,结合经验回放和目标网络,提升模型的学习效率与稳定性。与传统方法相比,此方法不但能提升带宽利用率,还能在保证关键数据流实时、可靠的情况下,让系统整体性能得到优化。

## 1 语义通信的模型框架

### 1.1 模型框架概述

本文构建了多个层次的语义通信模型框架,通过把语义通信与动态资源调度结合起来,从而优化带宽、功率和信道资源的分配。该框架由 5 个核心层次组成,如图 1 所示。(1)语义层要处理传输数据,做语义分析与编码,让系统能找出关键信息并分配优先级。该层会提取语音里的关键词、视频里的关键帧等信息,从而保证关键信息能优先传输。(2)传输层负责在网络里传输数据,还对带宽、功率、信道这些资源加以管理。传输层和语义层紧密耦合,依据语义优先级及网络状态调整资源分配策略,以此

保障高优先级数据流能高效传输。(3)感知层能感知网络环境与数据流的实时状态,给资源调度模型提供反馈信息,让智能体获取当前的带宽、信道质量等状况。(4)调度层依靠深度强化学习模型,动态调整带宽、功率分配等参数,优化资源分配。该层智能体在与环境交互中,逐渐掌握最优资源分配方法。(5)奖励函数依据语义恢复质量、传输延迟、资源消耗等因素进行计算,以此指导智能体优化资源调度。

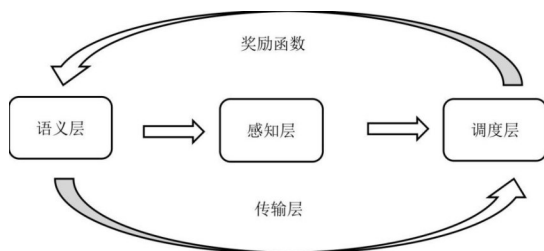


图1 基于深度强化学习的语义通信框架

## 1.2 语义优先级

在语义通信模型中,各数据流的语义优先级不同。为达成语义感知的资源调度目标,引入了数据流优先级模型。具体来说,依照数据流的实时性要求、语义重要性及带宽需求,给每个数据流分配一个优先级值。

实时性方面:因为语音、视频这类实时应用对数据延迟敏感度颇高,所以得优先调度。

语义重要性阐述:在实时视频传输中,关键帧对于复原画面质量具有核心作用,相比之下,非关键帧在带宽受限时可适当舍弃或进行压缩处理。

带宽需求:数据流在不同带宽条件下对传输质量有不同要求,系统依据当前带宽状况及数据流的具体需求进行动态调整。

根据上述要求,数据流的优先级规定如式(1)所示。

$$Q_i = w_1 R_i + w_2 I_i + w_3 B_i \quad (1)$$

式中: $Q_i$ 为优先级参数; $R_i$ 为数据流的实时性要求; $I_i$ 为数据流的语义重要性; $B_i$ 为数据流的带宽需求; $w_1$ 、 $w_2$ 、 $w_3$ 为权重系数,用于平衡各个因素在总优先

级中的影响。

依据优先级参数 $Q_i$ ,系统能够动态对资源分配予以调整,把更多带宽、功率之类的资源赋予优先级高的数据流,以此来提升整个系统的语义传输质量。

## 1.3 语义恢复与调度优化

语义恢复是语义通信中的关键环节,在语义通信系统中,通过数据冗余、前向纠错、语义压缩等方式来实现语义恢复。但是,这些方法通常会消耗非常多的网络资源,造成浪费,因此,本文提出了一种融合语义优先级与强化学习算法的动态资源调度框架来解决语义恢复效果与资源利用率之间的平衡问题。该框架能够依据语义恢复任务的优先级,对网络资源实施动态化的分配策略,从而在保障语义完整性的同时实现资源的高效利用。

在数学模型中,语义恢复能力 $S_i$ 被定义为数据流恢复的质量(如感知语音质量评估),并与奖励函数关联,如式(2)所示。

$$S_i = f(P_i, Q_i) \quad (2)$$

式中: $f$ 为映射函数,依据数据流的语义优先级和恢复质量来调整奖励; $P_i$ 为感知语音质量评估(perceptual evaluation of speech quality, PESQ)得分等度量,反映了每个数据流的语义恢复质量。

## 1.4 基于深度Q网络(DQN)的资源调度

调度层运用深度Q网络(DQN)去学习并优化资源分配策略。DQN代理会观察网络状态与语义优先级,从而选择合适的动作(如带宽分配、信道选择、功率控制等),还能根据所选动作的效果获取奖励。 $Q$ 值的更新规则如式(3)所示。

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)] \quad (3)$$

式中: $s_t$ 为当前状态, $a_t$ 为采取的动作, $r_t$ 为即时奖励, $\gamma$ 为折扣因子, $\alpha$ 为学习率。

该更新规则使得DQN通过学习在给定状态下可以获得动作的期望奖励,不断改进其策略,最终

收敛到最优的资源分配策略。传统 Q-learning 在高维状态空间下可能因数据相关性导致收敛困难,需通过经验回放和目标网络技术改进。

### 2 研究内容与方法

本文重点是基于深度强化学习的语义感知动态资源调度方法提高语义通信系统在复杂动态网络环境里的性能。针对语义通信系统里不同

优先级的数据流所具有的带宽需求,给出了基于深度强化学习的资源调度框架。在这个框架里,把数据流的语义优先级及网络的实时信道质量都整合到调度决策当中,从而达成动态带宽分配及低延迟传输的效果。深度 Q 网络(DQN)学习资源分配策略过程如图 2 所示。引入经验回放与目标网络技术,能让模型在复杂环境中学习更稳定。

算法 1: 基于深度 Q 网络 (DQN) 的动态资源调度融合算法
<p><b>输入:</b> 当前网络状态 <math>S_t=(B_t, C_t, D_t)</math>, 其中, <math>B_t</math> 为当前可用带宽, <math>C_t</math> 为信道质量(如信噪比), <math>D_t</math> 为经验回放缓冲区; 动作空间 <math>A=\{a_1, a_2, \dots, a_n\}</math>, 包括带宽分配、信道选择和功率控制。</p> <p><b>输出:</b> 最优资源调度策略 <math>\pi^*(S_t)</math>。</p> <p><b>过程:</b></p> <ol style="list-style-type: none"> <li>1. 初始化 Q 网络 <math>Q^*(S_t, a_t)</math>, 设置学习率 <math>\alpha</math>、折扣因子 <math>\gamma</math>、探索率 <math>\epsilon</math>。</li> <li>2. 从初始状态 <math>S_0</math> 开始, 执行以下步骤。               <ol style="list-style-type: none"> <li>2.1) 在当前状态 <math>S_t</math> 下, 选择一个动作 <math>a_t</math>, 使用 <math>\epsilon</math>-贪婪策略,                   <math display="block">a_t = \begin{cases} \operatorname{argmax}_a Q(S_t, a) &amp; \text{with probability}(1 - \epsilon) \\ \text{random action} &amp; \text{with probability } \epsilon \end{cases} .</math> </li> <li>2.2) 执行动作 <math>a_t</math>, 获取下一个状态 <math>S_{t+1}</math> 和奖励 <math>r_t</math>。</li> <li>2.3) 更新 Q 值。 <math>\theta \leftarrow \theta - \alpha \nabla \theta (r_{t+1} + \gamma \max_a Q(S_{t+1}, a'; \theta') - Q(S_t, a_t; \theta))^2</math> 。</li> <li>2.4) 更新状态 <math>S_t \leftarrow S_{t+1}</math>。</li> <li>2.5) 训练结束。训练达到最大步数或收敛时, 输出最优策略 <math>\pi^*(S_t)</math>。</li> </ol> </li> </ol>

图 2 DON 学习资源分配策略过程

#### 2.1 奖励函数

奖励函数应综合考虑带宽分配、延迟、语义优先级及信道质量,要保证优先级高的数据流能有更好的服务。在语义通信这一情形下,奖励函数把多个目标给封装好,从而让语义恢复最大化、延迟最小化、能耗降低、带宽利用率优化。奖励函数的定义如式(4)所示。

$$R_t = \alpha S_t - \beta L_t - \gamma E_t - \delta U_t \quad (4)$$

式中:  $\alpha, \beta, \gamma, \delta$  为平衡每个目标重要性的权重系数;  $S_t$  为语义信息恢复的质量,通过 PESQ 等指标来衡量;  $L_t$  为传输延迟;  $E_t$  为传输过程中消耗的能量;  $U_t$  为带宽的使用率。

通过适当调整权重系数,奖励函数可确保强化

学习代理优先考虑语义恢复,同时有效地管理延迟、能量和带宽。

#### 2.2 深度 Q 网络(DQN)的融合

DQN 算法在优化资源调度的收敛性方面以 Q-learning 原理为基础,通过双重网络(在线网络  $\theta$  与目标网络  $\theta'$ )分离动作选择与价值评估,避免 Q 值过估计问题。此外,经验回放机制通过随机采样历史数据优化策略,提升模型在动态网络环境中的适应性, Q 值函数的 Q-learning 更新规则  $Q(S_t, a_t; \theta)$  如式(5)所示。

$$Q(S_t, a_t; \theta) = E_{(s_{t+1}, r_{t+1}) \sim D} [r_{t+1} + \gamma \max_a Q(s_{t+1}, a'; \theta')] \quad (5)$$

状态空间 S 和操作空间 A 是有限的,奖励  $r_{t+1}$  是

有界的,折扣因子 $\gamma$ 范围满足 $0 \leq \gamma < 1$ , $D$ 为经验回放缓冲区,存储历史转移元组 $(s_t, a_t, r_{t+1}, s_{t+1})$ 的核心是通过目标网络 $\theta'$ 分离动作选择与价值评估,避免 $Q$ 值过估计。经验回放机制 $D$ 通过随机采样打破时序相关性,提升模型稳定性。利用经验重放和目标网络来稳定训练,使得DQN可以收敛到最优策略 $\pi^*(S_t)$ 。

### 2.3 语义优先级优化

首先把语义优先级纳入资源调度之中,然后将优化问题变为多目标框架的形式。要让目标函数达到最大值如式(6)所示。

$$\max_{a_t} E[R_t] = E \left[ \alpha \sum_{i=1}^n Q_i S_i - \beta L_t - \gamma E_t - \delta U_t \right] \quad (6)$$

合并语义优先级 $Q_i$ 到奖励函数里,强化学习代理能被激励,进而被激励分配更多资源以提升语义恢复,同时保障整体系统的效率。这种多目标优化,能保证即便在受限的网络环境里,也能把关键语义信息可靠地传输出去。在资源分配过程中,强化学习代理考虑到了各数据流的相对重要性。代理会优先考虑优先级较高的数据流,这样就能最大化所有数据流中语义恢复的加权和,从而在语义完整性与资源利用率之间达成最佳平衡。

## 3 实验设计及结果

### 3.1 实验设置

为衡量所提出的基于DQN的动态资源调度办法在语义通信里的有效性,在Windows 10操作系统下,使用Python 3.8,采用TensorFlow 2.4框架实现

DQN模型,使用OpenAI Gym框架进行环境封装与交互。实验运行平台的配置为Intel(R) Core(TM) i7-14650HX处理器、32 GB内存、NVIDIA GeForce RTX 4060显卡,确保模型训练与推理过程的稳定性和效率。实验设置学习率为0.001,折扣因子为0.99,经验回放缓冲区大小为10 000,目标网络更新频率为100步,模拟出3种资源分配的情形:(1)以固定带宽分配为基准,运用静态分配策略,不考虑网络状态及语义优先级;(2)传统强化学习分配(模拟SARSA算法)靠随机动作来达成动态分配,并未对语义优先级或信道质量加以整合;(3)基于DQN的语义感知调度,它把深度强化学习跟语义优先级及信道质量相结合,从而动态优化带宽分配策略,以满足高优先级流的需求。这种多维度的设计为资源分配策略在不同复杂性条件下的性能状况进行全面评估,能够对语义通信系统优化提供参照。实验设定了3种资源分配场景,具体参数如表1所示。

### 3.2 结果与分析

在本实验里,对固定带宽分配、传统强化学习分配及基于DQN的语义感知资源调度这3种资源分配办法展开了评估和比较。实验以带宽利用率、语义恢复质量(PESQ分数和关键帧准确性)这3个关键指标为基础,对不同方法在模拟网络环境下的性能作出全面剖析。以下将对各项指标逐一进行详尽阐述。

#### 3.2.1 带宽利用率

带宽利用率是衡量系统资源运用效率的关键

表1 场景参数

参数	固定带宽	传统强化学习	DQN
带宽分配方式	静态分配	随机分配	动态学习,优先级感知
动作空间	固定值	[0, 4]	[0, 4]
信道质量	无	无	有
语义优先级	无	无	有
强化学习机制	无	随机分配	经验回放+目标网络

指标,实验结果如图 3 所示。由图 3 可知,固定带宽被利用的占比达 70.00%,但静态分配策略无法动态适应网络流量需求,因此未能充分利用系统的带宽资源。传统强化学习方法通过优化经验来分配资源,使其利用率提高到了 72.44%。不过,它并没有结合语义优先级,使得带宽分配的效率提升有限。DQN 的语义感知资源调度的效果显著提升,利用率为 96.00%。该方法依据实时网络状况及语义优先级动态分配资源,有效提高了带宽利用率,充分发挥了系统资源的潜力。

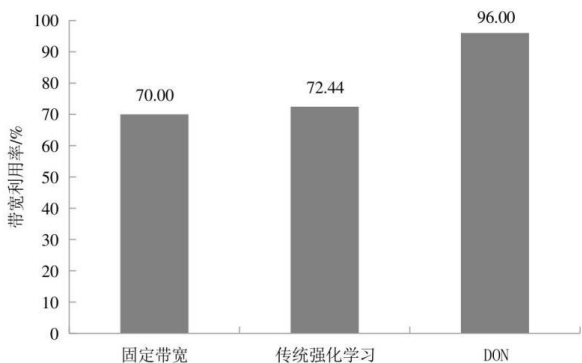


图 3 带宽利用率

### 3.2.2 语义恢复质量

语义恢复的质量是衡量语音流数据完整性和清晰度的重要指标,实验结果如图 4 所示。由图 4 可知,固定带宽分配方案所获得的 PESQ 分数是 3.02,满足了基本的语音通信质量要求,然而,该方案缺乏根据实时网络状况进行调整的能力。传统的强化学习方法的 PESQ 分数是 3.14,这比固定分配策略稍微高些。这就表明,它靠学习网络状态能在一定程度上提升语义恢复的质量。采用 DQN 方法后,PESQ 分数降低至 2.73。虽说该方法满足了高优先级的数据流需要,但可能在语音流的语义恢复中存在一定的性能牺牲,尤其是在带宽资源紧张的场景中。

### 3.2.3 传输延迟

传输延迟是评判系统响应能力的关键指标,实验结果如图 5 所示。由图 5 可知,固定带宽传输延

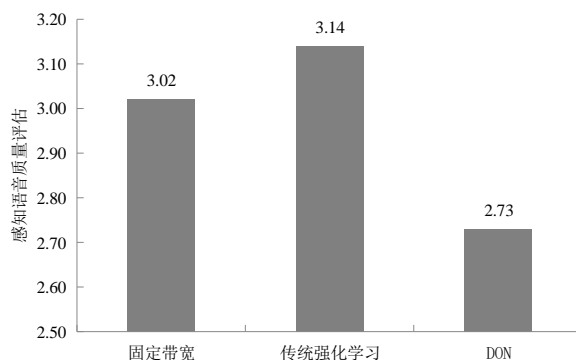


图 4 语义恢复质量

迟为 23.87 ms,其固定分配策略能保证基本的实时性,可是不能在动态环境下有效地调度资源;传统强化学习方法的传输延迟增加到 58.69 ms,这意味着其随机分配策略在高优先级流需求不满足时,会有一定程度的延迟;用 DQN 方法也做到了 8.98 ms 低传输延迟,这得益于其动态资源调度能力,能优先满足高优先级流的实时性需求,保证关键数据流能低延迟传输。

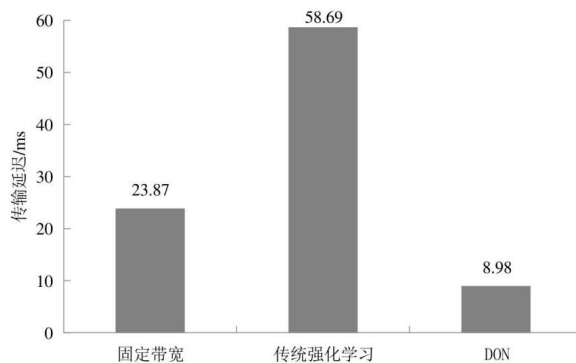


图 5 传输延迟

综上所述,在简单场景下,固定带宽分配的资源利用率较低,然而能耗与延迟均可控。传统强化学习的分配采用随机分配策略,在一定程度上提升了资源使用效率,可是,在高优先级流的保障方面表现不太理想。而基于 DQN 的语义感知调度,凭借动态优化能力以及语义优先级整合,能显著提升带宽利用率和高优先级流性能,不过,它在能耗上有一点增加,但整体表现依然是最优的。

## 4 结束语

本文提出了一种基于深度强化学习的语义感知动态资源调度框架,它把语义优先级和信道质量动态优化资源分配策略相结合。实验结果显示,这一方法不仅在带宽利用率、语义恢复质量、传输延迟控制等方面明显比固定带宽分配与传统强化学习方法要好,而且还能很好地在保障高优

先级流性能和提高系统资源利用效率之间达成平衡;虽然能耗增多了一些,然而性能的提高证明了语义感知调度是有效的。本文提出的框架为未来语义通信系统的资源优化提供了理论依据和实践指导。本文方法的主要局限在于能耗上升问题,后续研究可重点探索性能与能耗的平衡机制,同时进一步优化深度强化学习模型的训练效率和稳定性。

### 参考文献:

- [1] 唐伦,肖娇,赵国繁,等.基于能效的NOMA蜂窝车联网动态资源分配算法[J].电子与信息学报,2020,42(2):526-533.
- [2] 陈九九,郭彩丽,冯春燕,等.智能网联环境下面向语义通信的资源分配[J].物联网学报,2022(3):47-57.
- [3] WANG Y H, LI T H S, LIN C J. Backward Q-learning: the combination of sarsa algorithm and Q-learning[J]. Engineering Applications of Artificial Intelligence, 2013, 26(9): 2184-2193.
- [4] MNH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning [EB/OL]. (2013-12-19) [2025-01-12]. <https://arxiv.org/pdf/1312.5602>.
- [5] MNH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [6] SHANNON C E. A mathematical theory of communication[J]. Bell Systems Technical Journal, 1948, 27(4): 623-656.
- [7] 张平,牛凯,姚圣时,等.面向未来的语义通信:基本原理与实现方法[J].通信学报,2023(5):1-14.
- [8] 陈九九,冯春燕,郭彩丽,等.车联网中视频语义驱动的资源分配算法[J].通信学报,2021(7):1-11.
- [9] XU T, ZHAO M, YAO X, et al. An improved communication resource allocation strategy for wireless networks based on deep reinforcement learning[J]. Computer Communications, 2022, 188: 90-98.
- [10] WANG C, DENG D, XU L, et al. Resource scheduling based on deep reinforcement learning in UAV assisted emergency communication networks[J]. IEEE Transactions on Communications, 2022, 70(6): 3834-3848.
- [11] HAO Y, LI G Y, JUANG B F. Deep reinforcement learning based resource allocation for V2V communications[J]. IEEE Transactions on Vehicular Technology, 2019, 68(4): 3163-3173.
- [12] CHEN J Z, LIU W C, QUEVEDO D E, et al. Semantic-aware transmission scheduling: a monotonicity-driven deep reinforcement learning approach[J]. IEEE Communications Letters, 2023, 27(12): 3260-3264.

责任编辑:蒋召雪