基于D3QN的Wi-Fi网络智能调制方法

吴婷婷，方旭明
（西南交通大学信息科学与技术学院，四川 成都 611756）

摘　要：速率自适应（RA, rate adaptation）技术是 Wi-Fi 网络中的关键功能，能够根据实时观测的信道状态选择最优的数据传输速率。然而，现有的大多数速率自适应算法主要存在两个问题：一是依赖跨层信息反馈的方法在实际应用中较难实现；二是所采用的方法在速率选择策略上过于保守，当环境变化信噪比（SNR, signal-to-noise ratio）处于两个可选调制阶数之间时，均选择较低的速率阶数。为了解决这些问题，提出了一种基于深度强化学习中双决斗深度 Q 网络（D3QN, dueling double deep Q-network）的速率自适应算法，该算法无须进行跨层反馈，通过观测物理层信息来动态调整数据速率，并在奖励函数设计和模型加载阶段参考了现有的查表速率调节方法。仿真结果表明，所提算法相比其他 4 种基线方法，在不同场景中都能迅速适应环境变化，实现了更高的吞吐量性能。

关键词：Wi-Fi网络；深度强化学习；双决斗深度Q网络；自适应调制
中图分类号：TN929.5
文献标志码：A
doi：10.11959/j.issn.2096−3750.2025.00460

Intelligent modulation method for Wi-Fi networks based on D3QN

WU Tingting, FANG Xuming
School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China

Abstract: Rate adaptation (RA) technology is a key feature in Wi-Fi networks, capable of selecting the optimal data transmission rate based on real-time observed channel conditions. However, most existing rate adaptation algorithms exhibit two issues. Firstly, methods relying on cross-layer information feedback are often challenging to implement in practical applications. Secondly, the strategies employed are over-conservative in rate selection, opting for lower rates when the signal-to-noise ratio (SNR) varies between two selectable modulation levels. A rate adaptation algorithm based on the dueling double deep Q-network (D3QN) in deep reinforcement learning was proposed to address these issues. This algorithm eliminated the need for cross-layer feedback and dynamically adjusted the data rate through the observation of physical layer information. Additionally, it referenced existing table-based rate adjustment methods during the design of the reward function and model loading phase. Simulation results show that the proposed algorithm can rapidly adapt to environmental changes and achieve higher throughput performance compared with four baseline methods across various scenarios.

Key words: Wi-Fi networks, deep reinforcement learning, D3QN, adaptive modulation

0　引言

在无线通信网络中，信号传播会受到多种因素的影响，如路径损耗、阴影效应和外部干扰等，导致信道状态非常不稳定。为了更有效地利用无线资源，IEEE 802.11 无线局域网引入了速率自适应

收稿日期：2024−10−22；**修回日期：**2024−12−19
通信作者：方旭明，xmfang@swjtu.edu.cn
基金项目：国家自然科学基金资助项目（No. 62071393）；四川省重点研发计划项目（No. 2024YFHZ0093）
Foundation Items: The Natural Science Foundation of China (No. 62071393), Sichuan Science and Technology Program (No. 2024YFHZ0093)

（RA, rate adaptation）技术。RA技术能够根据当前无线信道的质量，动态选择合适的调制和编码方案（MCS, modulation and coding scheme）。由于信道条件存在波动，想要实现更高的吞吐量性能并不能单纯依赖于选择更高的物理速率。不同的MCS阶数都需要满足特定的信噪比（SNR, signal-to-noise ratio）要求，才能在接收端成功解码数据包[1]。因此，在信道状况良好时采用高MCS调制能有效地提升网络性能，但在较差的信道条件下使用高MCS阶数将导致较高的误码，从而降低网络性能。发送节点必须为数据包选择适当的MCS级别，尤其是在信道条件变化时，通过动态调整MCS阶数以实现高吞吐量。当前的IEEE 802.11网络采用了更多的天线、更宽的信道带宽和更高阶的调制，可用的MCS阶数显著增加[2]，速率自适应算法面临的挑战也随之增大。RA算法需要快速从众多选项中选择出在时变信道下可以实现最高吞吐性能的最佳速率，在信道变化较为频繁的Wi-Fi网络中，每种SNR条件下都有一个最优的传输速率，动态调整并选择最优的传输速率对于提升网络整体性能至关重要。

如今，机器学习技术已经被广泛应用于多个领域，并在提升Wi-Fi性能方面发挥着日益重要的作用。作为机器学习的一个分支，深度强化学习（DRL, deep reinforcement learning）擅长处理复杂环境中的优化问题[3-9]，在优化Wi-Fi网络的信道分配[10-11]、功率控制[12-14]、拥塞控制[15-17]等领域展现出了巨大的潜力。一方面，无线网络的环境具有高度复杂性和多变性，另一方面，目前实际采用的RA方法在速率选择策略上过于保守，当环境变化SNR处于两个可选调制阶数之间时，通常选择较低的速率阶数。因此，为了满足不同用户的需求，策略的选择必须具备足够的灵活性。在此背景下，本文旨在提出一种新型速率自适应算法，该算法利用深度强化学习模型，在无需任何跨层反馈的情况下，避免较为保守的低速率选择策略，通过物理层参数和奖励函数动态选择满足误码率要求的较高的调制速率，从而提升网络吞吐量性能。

1　相关研究

传统的IEEE 802.11网络速率自适应方案通常是基于规则的，主要分为基于SNR和基于采样的方法[18]。在基于SNR的方法中，发射端通过物理层估算SNR，并通过查表将其映射为当前信道条件下可支持的MCS阶数。而基于采样的RA算法通过多次探测每个MCS阶数，从中选择性能最佳的MCS阶数进行数据传输。这些传统方法在应对快速变化的无线环境时通常存在响应不及时的问题，为了解决这一问题，基于机器学习的RA算法应运而生，显著地提升了网络性能。文献[18-23]中，研究者们探索了基于机器学习的速率自适应算法。文献[18]针对采样算法的性能瓶颈，提出了一种基于神经网络的RA算法NeuRA，NeuRA通过预测不同速率的预期吞吐量，显著地提升了采样效率，减少了帧丢失并提高了帧聚合算法的运行效率。文献[19]提出了基于机器学习驱动的MLRA算法，MLRA算法利用两阶段操作筛选出可能最大化吞吐量的速率，并对这些速率进行精细搜索，确定出最佳速率。文献[20]所提的SARA算法将速率自适应问题视为一个学习自动机，自动机通过与随机的无线信道环境交互，逐步学习最佳传输速率，并依据概率选择合适的数据速率。文献[21]运用神经网络来学习成功和失败阈值与争用情况之间的关系，通过表查找快速切换速率，依据当前争用情况选择最优的成功和失败阈值。此外，文献[22]指出，在直线传播场景中使用精细定时测量功能选择速率能够减少误差，通过MCS、SNR和成功传输概率的关系可以选择出最合适的速率。文献[23]提出了一种基于随机多臂老虎机的分布式算法，通过探索不同配置（如信道带宽和MCS阶数）对网络性能的影响，以帧成功率作为奖励进行调整。

随着深度强化学习的迅速发展，基于DRL的RA方案逐渐成为研究热点。文献[2]提出的drlRA算法通过考虑随机信道访问引发的冲突影响MCS选择策略，将等待时间纳入到奖励函数的设计，并且对由MCS决策引起的数据包错误和由冲突引起的数据包错误进行区分，实现了较高的总体吞吐量。文献[3]提出了一种针对Wi-Fi网络应用层速率的自适应DRL模型，该模型基于当前的网络性能，在奖励函数中考虑了体验质量，从而在网络拥塞时实现了更优的吞吐量性能。文献[24]提出的DARA算法通过对发射机接收帧的观察来动态调整MCS阶数，使用基于轨迹的模拟克服了传统RA算法调整速度较慢的问题。文献[25]提出了一种基于三维迷宫的RA算法，将MCS阶数、多输入多输出模式

和带宽等速率相关特征抽象为三维空间中的坐标系，通过在这个空间中探索路径，找到具有最佳有效吞吐量的速率。文献[26]提出了一种具有双时间尺度的深度Q网络（DQN, deep Q-network）算法对码本选择和MCS自适应方案进行联合优化，该算法将联合问题公式化为一个双时间尺度系统，在小时间尺度上执行MCS自适应，在大时间尺度上优化码本。通过闭环决策指导和奖励评估，进一步修改传统的DRL训练机制，获得了更高的链路数据速率。文献[27]提出了一种智能方法，通过选择合适的调制编码方案来最大化系统的编码率。该方法采用DQN算法动态选择MCS，结合外环链路自适应算法来增强MCS的选择过程，并提出了一种创新的训练机制，能够在训练阶段提升DQN模型的性能和收敛速度，从而有效地提高编码率。

综上所述，传统的速率自适应方案因简单、快捷的优点被广泛应用。然而，当环境发生快速变化，需要频繁调整速率时，基于查表规则的方案表现得不够灵活。当SNR位于两个可选调制阶数之间时，算法通常倾向于选择较低的速率阶数，导致系统性能无法达到最优水平。而引入DRL的智能方案则能有效地解决这一问题，代理通过实时统计网络的传输数据情况，依据传输成功率、丢包率等关键指标动态地调整传输速率，从而快速地响应环境的变化，提升系统性能。然而，这些经典的基于DRL的速率自适应算法在实际应用中存在两个局限性。首先，由于部分算法使用的机器学习模型难以收敛或收敛速度较慢，速率在不同状态之间频繁波动，并且在提升速率时比降低速率更保守。因此，当网络环境恶化时，传输速率通常处于下降状态，网络性能难以达到最佳水平。其次，实际应用的速率调整框架通常依赖预定义的查找表，将接收的SNR映射到特定的MCS阶数，而部分依赖跨层信息（如丢包率和确认信号反馈）的DRL速率自适应算法在实际应用中难以有效实现。因此，迫切需要一种能够根据环境信息，尽可能选择满足误码率要求下较高MCS阶数的自适应速率选择框架，不依赖跨层反馈，满足实际应用的需求。

2　基于D3QN的RA方法

2.1　系统模型

本文研究了一个基于SNR状态输出MCS动作的Wi-Fi网络速率调整系统，系统场景如图1所示。该系统包含一个实现IEEE 802.11ax标准的发射器AP和多个接收器STA，不考虑其他干扰因素。其中，只有一个与AP相关联的目标接收器STA使用本文所提的速率调整算法。本文的目标是在发射器上训练一个代理，该代理基于前一个观察周期t_{n-1}内接收帧的SNR及相邻两个观察周期内的SNR变化，学习在下一观察周期t_n内发送至目标STA的数据帧传输的最佳MCS阶数，从而不受移动模式或通信标准的影响。

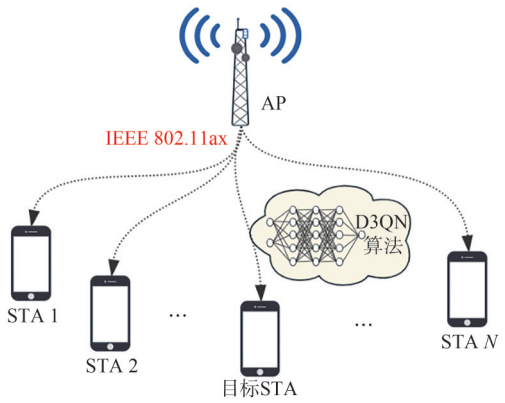

图1　系统场景

将代理感知到的信道环境状态看作一些离散的时间序列，整个速率控制过程就可以建模为一个马尔可夫决策过程（MDP, Markov decision process）。强化学习任务可以通过MDP五元组$\langle S, A, R, P, \pi\rangle$进行描述：智能体的所有状态$s_t$构成状态集合$S$，在每个状态下所有动作$a_t$组成动作集合$A$，在各状态下所有动作获得的反馈奖励$r_t$组成奖励函数$R: S \times A \to R$，状态之间的转移概率为$P: S \times A \xrightarrow{P} R$。在每个时刻，智能体根据所处的状态以一定的概率选择相应动作，这个由状态到动作的映射过程称为智能体的策略$\pi: S \to A$。为了评估在当前策略下某一状态的优劣，引入状态价值函数，其定义为当前状态对后续累积回报的影响，表示为

$$v_\pi(s) = \mathrm{E}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t = s\right] \tag{1}$$

其中，r_{t+k}表示某个状态下的奖励，γ表示折扣因子，取值范围为[0,1]。

2.2　状态和动作空间

代理在一个观察周期t内的观察信息s_t包含信道状态和信道变化信息，可以表示为

$$s_t = \{\mathrm{SNR}_t, \Delta\mathrm{SNR}_t\} \tag{2}$$

其中，SNR_t代表一个观察周期内的平均SNR，通过对当前观察周期内接收到的所有数据包的SNR取平均值得到。$\Delta\mathrm{SNR}_t=|\mathrm{SNR}_t-\mathrm{SNR}_{t-1}|$，表示观察周期$t$相较于上一观察周期$t-1$内平均SNR的变化值。具体来说，SNR反映了当前的链路质量，SNR变化值反映了链路状态的动态变化趋势。

根据Wi-Fi标准，速率选择取决于物理层的MCS方案，因此，将动作空间A定义为IEEE 802.11ax标准中可用的MCS速率索引值，包含12种可能的操作（索引值0~11分别对应MCS0-MCS11）[28]。在每个观察周期内保持选择的动作不变，直到下一个观察周期输出新的动作。

2.3　奖励函数

为代理设计合理的奖励机制对于学习出一个有效的策略至关重要。当环境变化SNR处于两个可选调制阶数之间时，传统方案均选择较低的速率阶数，致整体性能下降。为了避免这种较为保守的策略，需要确保在对MCS阶数进行调整时，代理在满足误码率要求的前提下，尽可能选择更高级别的MCS阶数，从而实现更高的吞吐量性能。为了确保尽可能高的MCS和实际传输速率，本文在奖励函数的设计中同时考虑了这两个指标。

在状态s_t下动作的奖励r_t定义为

$$r_t = \begin{cases} \alpha\cdot\dfrac{\mathrm{Thr}_t}{\mathrm{Thr}_{\mathrm{ratetable}}} + (1-\alpha)\cdot\dfrac{\mathrm{MCS}_t}{\mathrm{MCS}_{\max}}, \\ \quad t\text{时间内目标STA传输} \\ 0 \quad , \quad \text{其他} \end{cases} \tag{3}$$

其中，α为权重因子，Thr_t为实际的物理层传输速率，定义为网络中节点物理层收到的比特总数与传输时间的比率。$\mathrm{Thr}_{\mathrm{ratetable}}$是根据当前SNR和预设比特误码率（BER, bit error rate）阈值在IEEE P802.11 TGax[29]提供的BER-SNR表中查找到的传输速率，此表与现有实际应用的速率调整方法参考的速率表相同。为了鼓励代理追求高MCS速率，奖励的另一部分表示为当前动作MCS_t与最高MCS速率$\mathrm{MCS}_{\max}$的比值。只有当STA进行传输时才给予代理积极的奖励，奖励函数鼓励代理学习能够获得较高吞吐量的MCS速率。

2.4　基于D3QN的RA方法

在本文的模型中，环境观测信息包括SNR的变化值，因此，所选择的MCS动作并不能完全准确地决定下一个状态（SNR差值）的转移。此外，由于状态空间包含连续变量，不同状态的数量非常庞大，这种情况适合使用参数化的动作价值函数来选择最优的动作。目前，较为先进的RA算法[2, 24]均采用DQN的方法实现更高的总体吞吐量。然而，DQN算法根据贪婪算法计算Q值，导致对目标Q值的过高估计，使最终得到的算法模型存在很大的偏差，而更先进的双重深度Q网络（DDQN, double deep Q-network）[30]算法在速率调整问题上较难收敛，但双决斗深度Q网络（D3QN, dueling double deep Q-network）[3, 31]算法通过优势函数直接学习状态的价值，使在某些动作不直接影响环境的情况下，算法能够快速学习到最优策略。D3QN算法相较于DQN算法新增了DDQ和对抗深度Q网络（dueling DQN）技术，DDQN通过引入两个神经网络分别选择动作和计算目标Q值，从而减少过高估计的影响。Dueling DQN则将Q值分解为状态价值函数和优势函数，帮助模型更有效地估计Q值。基于上述原因，本文采用D3QN算法来选择最优MCS阶数。

D3QN算法模型如图2所示，算法包含两个结构完全相同的神经网络，分别为Q网络和目标网络。网络的输入为环境状态信息，输出为根据环境状态信息可以执行的各种动作对应的Q值$Q(s, \cdot; \theta)$，其中，θ代表网络的权重。

代理决策时采用ε-greedy策略以ε的概率随机选择一个动作（MCS阶数），以$1-\varepsilon$的概率选择Q网络计算得到的最大Q值对应的动作，如式(4)所示。

$$a_t = \begin{cases} a_i, i = (0, 1, \cdots, 10, 11), \text{概率为}\varepsilon \\ \arg\max Q(s_t, a_t; \theta_t), \text{概率为} 1-\varepsilon \end{cases} \tag{4}$$

在D3QN算法中，神经网络的输出$Q(s, a; \theta_t)$由当前状态的状态价值函数Value和每个动作的优势函数Advantage组成，采用这种对偶网络结构能够更好地区分状态价值和动作优势，从而提高策略评估的准确性。优势函数Advantage计算式如式(5)所示。

$$A(s, a; \theta_t) = Q(s, a; \theta_t) - V(s; \theta_t) \tag{5}$$

其中，θ_t代表Q网络的权重参数，$A(s, a; \theta_t)$表示动作a相对于$V(s; \theta_t)$的优势，选择的动作a越好，优势值越大。$V(s; \theta_t)$表示未来步骤中采取概率行动的总体预期回报，用于评价状态s的好坏，计算式如式(6)所示。

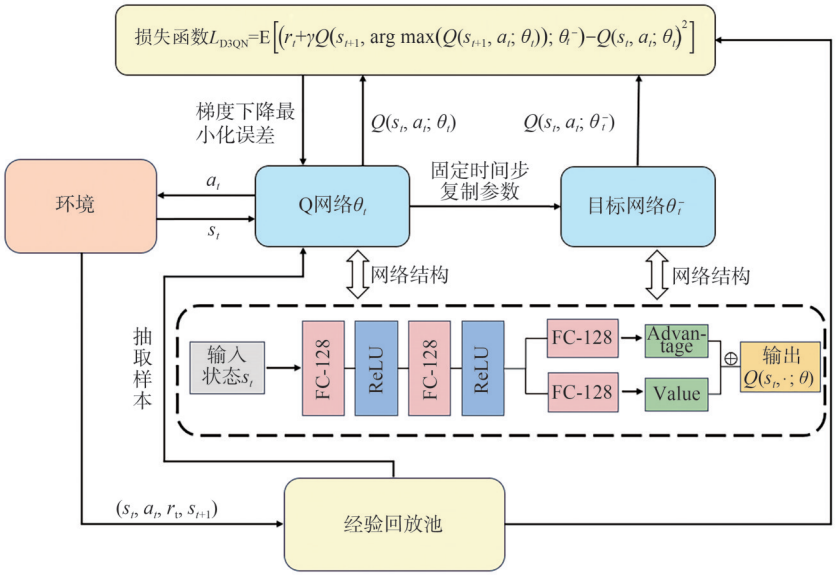

图2　D3QN 算法模型

$$V(s;\theta_t)=\max_{\pi} V(s;\theta_t) \tag{6}$$

引入优势函数$A(s, a; \theta_t)$是为了表示特定动作相对于其他动作的优劣，对于给定的状态s，如果所有动作的优势值$A(s, a; \theta_t)$都相同，将无法真正反映动作间的差异。为了消除每个动作优势值的偏置，需要从Q值中去掉平均优势值，最终得到的Q值$Q(s, a; \theta_t)$如式(7)所示。

$$Q(s,a;\theta_t)=V(s;\theta_t)+\left(A(s,a;\theta_t)-\frac{1}{|A|}\sum A(s,a;\theta_t)\right) \tag{7}$$

D3QN 算法使用损失函数更新神经网络权重参数，计算式如式(8)所示。

$$L_{D3QN}=E\left[\left(r_t+\gamma Q\left(s_{t+1}, \arg\max\left(Q(s_{t+1}, a_t; \theta_t)\right); \theta_t^-\right)-Q(s_t, a_t; \theta_t)\right)^2\right] \tag{8}$$

其中，θ_t^-代表目标网络的权重参数，r_t为奖励函数，γ为折扣因子。D3QN 算法与Q学习算法[32]相同，将单步奖励值r_t和状态转移后可能取得的最大折扣Q值$\gamma Q(s_{t+1}, \arg\max(Q(s_{t+1}, a_t; \theta_t)); \theta_t^-)$作为算法每一步更新的目标。当$\gamma$值为0时，表示代理只关注当前奖励，采取短期策略；$\gamma$值越高，表示代理越关注未来状态−动作对获得的收益，倾向于采取长期策略。$Q(s_t, a_t; \theta_t)$为Q网络输出的预测值，采用均方差函数，进行误差反向传播。

具体来说，代理在当前状态下会选择Q网络中Q值最大的动作，并通过目标网络计算该动作对应的Q值。每当Q网络经过N次迭代训练后，将网络中的参数复制到目标网络中，进行更新。

基于D3QN的MCS速率优化算法如算法1所示，总结了基于D3QN的MCS速率优化算法的伪代码。为了充分学习到最优策略，本文将智能体与环境交互产生的一系列经验序列(s_t, a_t, r_t, s_{t+1})存储在经验回放池中作为训练样本，每次训练时从中抽取一定批量大小的经验序列。随着训练步数的增加，经验回放池中的样本也越来越丰富，当信道状态变化缓慢时，采用经验回放机制能够打破连续经验之间的强相关性，使每次训练更具独立性，结果更可靠[19]。

算法1　基于D3QN的MCS速率优化算法

初始化　对Q网络和目标网络进行初始化，对经验回放池P和训练池的容量进行初始化，定义最大训练步数N、批量大小b；训练步数n=0；初始化信道状态

while $n<N$ **do**

　获取状态s作为输入；

　根据ε−greedy策略选择动作；

　执行动作a_t，获得奖励r_t和新的环境状态s_{t+1}；

　if 训练阶段 **then**

　　将经验转移序列(s_t, a_t, r_t, s_{t+1})存储到经验回放池中；

　　if $P>b$ **then**

　　从经验回放池中获得数量为b的经验序列；

　　计算损失函数L；

　　通过梯度下降法更新Q网络参数θ_t；

更新目标网络参数θ_t^-；

更新参数ε；

end if

end if

更新状态和动作；

end while

2.5　模型运行

2.5.1　训练阶段

训练时贪婪因子ε从1开始，ε的值随训练迭代次数的增加而减小，当训练到一定的迭代次数时，ε的值减小到0.1并保持不变。这样可以保证在训练过程中代理能够充分探索状态和动作空间，提高训练的稳定性。

2.5.2　评估阶段

在数据包传输过程中，SNR会对不同MCS的表现产生影响。针对每种MCS，当SNR变化时，传输成功率也会发生变化，其变化幅度与当前SNR有关。因此，可以依据实时SNR值大致地估计出最佳传输速率可能的范围。然而，索引相邻的MCS所需的SNR值差异较小，单凭SNR值可能不足以明确判断哪个速率的传输成功率更高。在这种情况下，可以参考预定义的BER-SNR表判断当前的SNR是否达到SNR阈值，如果达到了MCS调节的SNR阈值，则可以使用基于D3QN的RA算法更细化地选择最佳速率进行传输。另外，注意到实际基于查表模式的自适应MCS算法的SNR调节阈值基本相差几分贝，因此同样可以根据SNR变化差值使用基于D3QN的RA算法细化速率选择。

当环境状态达到预设的SNR阈值或相邻时刻SNR差值较大时，加载训练好的模型以触发速率自适应，否则将保持当前的MCS速率进行传输，这种方法不仅可以确保速率选择的平稳性，还能便于实际应用。其中，SNR阈值根据IEEE P802.11 TGax提供的BER-SNR表选取BER低于10%时映射的SNR值得到。在评估阶段，将贪婪因子ε设置为0，避免代理进行探索尝试，并停止经验回放机制，使MCS速率的选择完全依赖于训练好的模型。模型运行流程如图3所示。

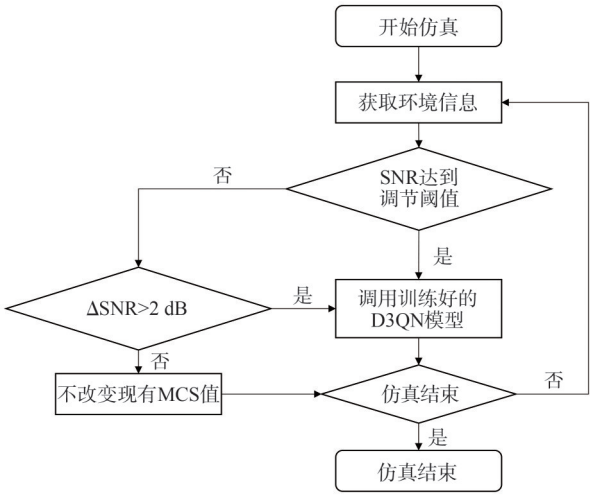

图3　模型运行流程

3　仿真测试

3.1　参数设置

本文采用NS-3作为仿真平台进行网络场景搭建，并使用Pytorch库实现D3QN算法，以评估MCS速率选择算法的性能。实验在Linux环境中进行，使用NS-3（版本3.37）和ns3-ai接口进行开发，网络的无线传输路径损耗评估、错误率计算及节点移动行为均采用NS-3内置模型实现。ns3-ai是对ns3-gym的改进，实现了内存共享并为基于Python的AI框架和NS-3之间提供了高效、高速的数据交换[33]。NS-3仿真采用的主要参数见表1。

表1　　NS-3仿真采用的主要参数

仿真参数	设定值
Wi-Fi标准	IEEE 802.11ax
信道带宽	80 MHz
AP发射功率	20 dBm
工作频率	5 GHz
流量类型	UDP
数据包大小	1 500 Byte
训练时长	50 s/150 s
路径损耗模型	Log-distance
错误率模型	TableBasedErrorRateModel
移动模型	ConstantVelocityMobilityModel
移动速度	均匀随机变量（4~6 m/s）

3.2　评价指标

本文选择目标STA的MAC层吞吐量作为评估不同速率控制算法性能的指标。将所提的基于D3QN的RA算法与4种经典速率控制算法进行比较，包括MinstrelHT算法[34]、Thompson采样算法、基于DQN的RA算法以及基于DDQN的RA算法。MinstrelHT算法是Linux内核中默认的速率适配算法，通过跟踪每个可用MCS速率成功发送帧的概率，将概率乘以速率计算出预期吞吐量并据此选择最佳的MCS。Thompson采样算法是NS-3中提供的

基于机器学习的速率选择方法。D3QN 算法主要参数设置见表 2。

表 2　　D3QN 算法主要参数设置

仿真参数	设定值
经验回放池容量	100 000
贪婪因子 ε	[0.1,1.0]
目标网络更新频率	每训练 500 步
隐藏层维度	128
批量大小	128
学习率	0.001
折扣因子 γ	0.95
折扣因子 α	0.80
损失函数	均方误差

3.3　仿真结果

为了评估所提算法在静态场景中的性能，本文设置了几组处于不同 SNR 下的静态环境，对于每次评估，取每个 RA 算法运行 5 次的平均吞吐量作为结果。其中，STA 保持静止并且与 AP 的距离固定，网络中存在 N 个 STA，以 N=10 为例。训练的总持续时间设置为 50 s，评估阶段时间设置为 30 s，算法预热时间为 5 s。需要说明的是，本文训练时的软件、硬件配置为：Windows 11 操作系统，Intel i5-12500H@2.50 GHz 处理器，内存为 16 GB，12 内核，中央处理器的双精度浮点运算能力约为 480 GFLOPS。在实际应用中，为了适应快速变化的信道状态，需要缩短算法的训练时间，若要将训练时间降低到 5 s，则需要将设备的运算能力提高约 10 倍。总体来说，为了实现算法在实际中的应用，通常需要依赖高运算力的设备支持。5 种算法在不同 SNR 下的性能表现如图 4 所示。

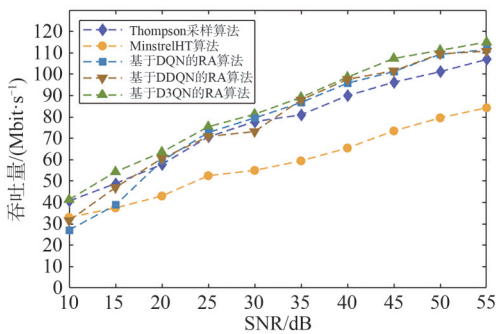

图 4　5 种算法在不同 SNR 下的性能表现

在图 4 中，采用基于 D3QN 的 RA 算法，代理充分观察环境后选择最佳 MCS 阶数，因此，STA 的吞吐量随着 SNR 的增大而呈现规律性的增加，采用其他算法的性能也是如此。在不同静态环境中，MinstrelHT 算法基于保守策略选择了较低的 MCS，性能表现较差，而基于 D3QN 的 RA 算法和 Thompson 采样算法能够很好地适应信道环境，选择更高的 MCS，从而获得了较高的吞吐量。基于 DQN 的 RA 算法和基于 DDQN 的 RA 算法在较低的 SNR 条件下吞吐量明显低于其他 3 种算法，主要原因在于代理无法有效地区分相似状态之间的差异。在低 SNR 环境中，代理错误地选择了较高的 MCS 阶数，导致接收端无法成功解码数据包。从结果可以得出结论，在不同的 SNR 下，所提的基于 D3QN 的 RA 算法都可以根据过去积累的经验智能地调整 MCS，实现对速率的平稳调整，并且对比其他速率调整算法实现了更高的吞吐量。与 MinstrelHT 算法相比，所提算法实现了 25.25%~50.59% 的吞吐量增益，与 Thompson 采样算法相比，所提算法实现了 1.69%~11.57% 的吞吐量增益，与基于 DQN 的 RA 算法相比，所提算法实现了 1.76%~52.68% 的吞吐量增益，与基于 DDQN 的 RA 算法相比，所提算法实现了 1.03%~31.09% 的吞吐量增益。

为了评估时变环境中不同速率自适应算法的性能，本文考虑这样一个移动场景，网络中存在 1 个 AP 和 n 个 STA，n=1。STA 节点以 4~6 m/s 的速度远离 AP 节点，直至运动到距离 AP 节点 100 m 处时再次靠近 AP 节点，如此循环往复，模拟时变场景。模型训练阶段的总持续时间设置为 150 s，评估阶段时间为完成一次往返运动的时长，在此期间，将 STA 每运动 20 m 得到的平均吞吐量作为性能指标，5 种算法在 STA 逐渐远离 AP 场景中的性能表现如图 5 所示，5 种算法在 STA 逐渐靠近 AP 场景中的性能表现如图 6 所示。

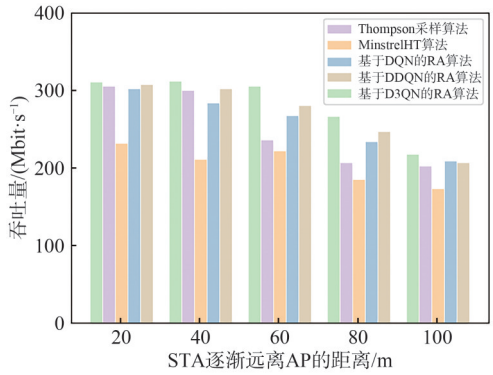

图 5　5 种算法在 STA 逐渐远离 AP 场景中的性能表现

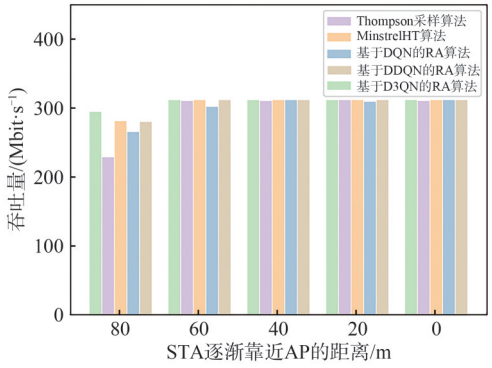

图6　5种算法在STA逐渐靠近AP场景中的性能表现

从STA逐渐远离AP的结果来看，基于D3QN的RA算法的表现始终优于其他4种算法，算法实现的吞吐量比MinstrelHT算法高出25.28%~47.68%，比Thompson采样算法高出1.80%~29.45%，比基于DQN的RA算法高出2.74%~14.46%，比基于DDQN的RA算法高出1.24%~9.07%。MinstrelHT算法在时变环境中无法区分不同的信道环境，从而选择了不准确的MCS，导致STA的吞吐量出现了部分波动。Thompson采样算法在逐渐到达60 m距离时性能严重下降，而基于DRL的3种RA算法在距离AP较远时仍能保持较高的吞吐量性能。基于D3QN的RA算法可以更快地跟踪信道环境的变化，因为它可以根据经验丰富的DRL模型避免选择保守的较低的MCS，在逐渐运动到最远处时依旧能选择合适的高MCS，实现更高、更稳定的吞吐量。从STA逐渐靠近AP的结果来看，基于D3QN的RA算法依然比其他4种算法更快地找到最优速率，算法实现的吞吐量比MinstrelHT算法高出0.02%~4.86%，比Thompson采样算法高出0.39%~28.41%，比基于DQN的RA算法高出0.02%~11.16%，比基于DDQN的RA算法高出0.02%~5.29%。

为了探究学习率大小对模型训练的影响，本文对比了不同学习率下的训练结果。以SNR=40 dB的静态场景为例，训练的总持续时间仍为50 s，网络中存在N=10个STA，其他参数设置如表2所示。每隔500 ms统计一次目标STA的吞吐量，不同学习率下的训练结果如图7所示。学习率对模型训练有较大的影响，学习率过大或过小都不利于模型快速地收敛到最优解。较大的学习率会导致模型权重更新过快，无法有效地捕捉不同状态之间的差异，虽然能快速地达到较高的吞吐量值，但最终导致模型发散。相反，学习率过小可能导致模型陷入局部最小值，无法找到最优策略。当学习率设置为0.001时，模型能够充分学习，吞吐量逐渐趋于平稳，表示代理已经成功学习到适应环境的最优策略。

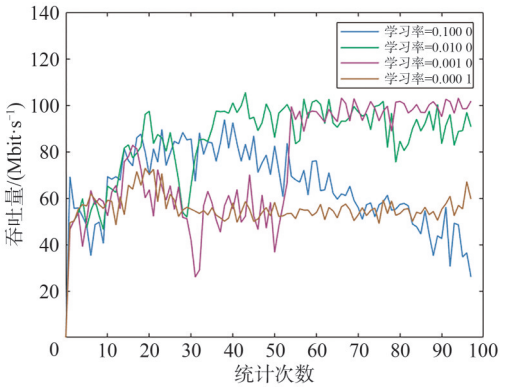

图7　不同学习率下的训练结果

此外，为了探究损失函数式(8)中折扣因子γ对模型训练的影响，本文比较了D3QN算法在不同γ取值下的训练结果。仍然以SNR=40 dB的静态场景为例，每隔500 ms统计一次目标STA的吞吐量，训练总时长为50 s，其他参数设置如表2所示，不同γ取值下的训练结果如图8所示。

在图8中，较低的γ值使代理更关注当前动作的奖励，而忽视了可能带来长远收益的动作，从而无法学习到最优策略；当γ值过大时，代理的策略过于偏向未来收益，由于当前动作与未来状态的联系较弱，同样导致无法学习到最优策略。当γ值设置为0.95时，能够有效地平衡当前奖励与未来收益的比重，使代理能够选择最佳动作并实现较快的收敛。

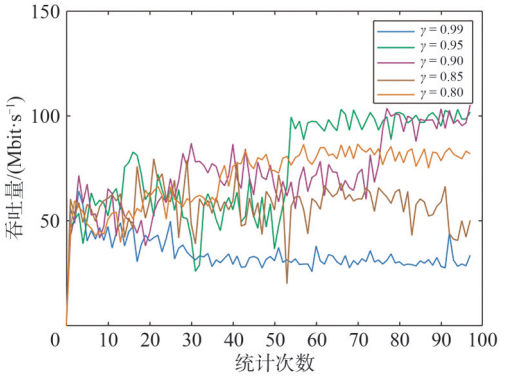

图8　不同γ取值下的训练结果

4　结束语

本文介绍了一种基于D3QN算法并适用于Wi-Fi网络的速率自适应方法。所提方法依赖物理层状态

信息选择最佳MCS速率，避免选择较保守的低速率策略，为了更好地实现模型，将在线学习和传统方法结合，在奖励函数和加载模型阶段参考了实际常用的查表速率调节方法，以便于应用实现。本文的研究结果表明，在不同的信道场景中，所提方法与先进的同类算法相比实现了更好的吞吐量性能。此外，计划进一步研究多参数联合优化模型，具体包括联合调整RA算法与其他物理层参数。

参考文献：

[1] PESERICO G, FEDULLO T, MORATO A, et al. SNR-based reinforcement learning rate adaptation for time critical Wi-Fi networks: assessment through a calibrated simulator[C]//Proceedings of the 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). Piscataway: IEEE Press, 2021: 1-6.

[2] LIN W H, GUO Z Y, LIU P, et al. Deep reinforcement learning based rate adaptation for Wi-Fi networks[C]//Proceedings of the 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall). Piscataway: IEEE Press, 2022: 1-5.

[3] SAMMOUR I, CHALHOUB G. Application-level data rate adaptation in Wi-Fi networks using deep reinforcement learning[C]//Proceedings of the 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall). Piscataway: IEEE Press, 2022: 1-7.

[4] LI Z R, WANG X, PAN L, et al. Network topology optimization *via* deep reinforcement learning[J]. IEEE Transactions on Communications, 2023, 71(5): 2847-2859.

[5] CALLEGARO D, LEVORATO M, RESTUCCIA F. SmartDet: context-aware dynamic control of edge task offloading for mobile object detection[C]//Proceedings of the 2022 IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM). Piscataway: IEEE Press, 2022: 357-366.

[6] QIN P, FU Y, ZHANG J, et al. DRL-based resource allocation and trajectory planning for NOMA-enabled multi-UAV collaborative caching 6G network[J]. IEEE Transactions on Vehicular Technology, 2024, 73(6): 8750-8764.

[7] AL-ERYANI Y, HOSSAIN E. Self-organizing mmWave MIMO cell-free networks with hybrid beamforming: a hierarchical DRL-based design[J]. IEEE Transactions on Communications, 2022, 70(5): 3169-3185.

[8] HUANG R, WONG V W S. Joint user scheduling, phase shift control, and beamforming optimization in intelligent reflecting surface-aided systems[J]. IEEE Transactions on Wireless Communications, 2022, 21(9): 7521-7535.

[9] KURMA S, KATWE M, SINGH K, et al. Spectral-energy efficient resource allocation in RIS-aided FD-MIMO systems[J]. IEEE Transactions on Wireless Communications, 2024, 23(5): 5125-5141.

[10] ZHONG Y, CHEN H, LIU W, et al. Deep reinforcement learning based channel allocation for channel bonding Wi-Fi networks[C]//Proceedings of the 2023 19th International Conference on Mobility, Sensing and Networking (MSN). Piscataway: IEEE Press, 2023: 113-119.

[11] CHEN H, LIU P, YOU L Z, et al. Deep reinforcement learning based dynamic channel bonding for Wi-Fi networks[C]//Proceedings of the 2023 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom). Piscataway: IEEE Press, 2023: 153-160.

[12] HUANG Y W, CHIN K W. A hierarchical deep learning approach for optimizing CCA threshold and transmit power in Wi-Fi networks[J]. IEEE Transactions on Cognitive Communications and Networking, 2023, 9(5): 1296-1307.

[13] EL JAMOUS Z, DAVASLIOGLU K, SAGDUYU Y E. Deep reinforcement learning for power control in next-generation WiFi network systems[C]//Proceedings of the MILCOM 2022 - 2022 IEEE Military Communications Conference (MILCOM). Piscataway: IEEE Press, 2022: 547-552.

[14] LUO Y Z, CHIN K W. An energy efficient channel bonding and transmit power control approach for WiFi networks[J]. IEEE Transactions on Vehicular Technology, 2021, 70(8): 8251-8263.

[15] WYDMANSKI W, SZOTT S. Contention window optimization in IEEE 802.11ax networks with deep reinforcement learning[C]//Proceedings of the 2021 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2021: 1-6.

[16] GRASSO C, RAFTOPOULOS R, SCHEMBRA G. OSCAR: a contention window optimization approach using deep reinforcement learning[C]//Proceedings of the ICC 2023 - IEEE International Conference on Communications. Piscataway: IEEE Press, 2023: 459-465.

[17] PURWITA A A, YESILKAYA A, HAAS H. Intelligent subflow steering in MPTCP-based hybrid Wi-Fi and LiFi networks using model-augmented DRL[C]//Proceedings of the GLOBECOM 2022-2022 IEEE Global Communications Conference. Piscataway: IEEE Press, 2022: 425-430.

[18] KHASTOO S, BRECHT T, ABEDI A. NeuRA: using neural networks to improve WiFi rate adaptation[C]//Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems. New York: ACM Press, 2020: 161-170.

[19] LI C Y, CHEN S C, KUO C T, et al. Practical machine learning-based rate adaptation solution for Wi-Fi NICs: IEEE 802.11ac as a case study[J]. IEEE Transactions on Vehicular Technology, 2020, 69(9): 10264-10277.

[20] JOSHI T, AHUJA D, SINGH D, et al. SARA: stochastic automata rate adaptation for IEEE 802.11 networks[J]. IEEE Transactions on Parallel and Distributed Systems, 2008, 19(11): 1579-1590.

[21] WANG C. Dynamic ARF for throughput improvement in 802.11 WLAN via a machine-learning approach[J]. Journal of Network and Computer Applications, 2013, 36(2): 667-676.

[22] CIEZOBKA W, WOJNAR M, KOSEK-SZOTT K, et al. FTM-Rate: collision-immune distance-based data rate selection for IEEE 802.11 networks[C]//Proceedings of the 2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM). Piscataway: IEEE Press, 2023: 242-251.

[23] KARMAKAR R, CHATTOPADHYAY S, CHAKRABORTY S. Dynamic link adaptation in IEEE 802.11ac: a distributed learning based approach[C]//Proceedings of the 2016 IEEE 41st Conference on Local Computer Networks (LCN). Piscataway: IEEE Press, 2016: 87-94.

[24] QUEIRÓS R, ALMEIDA E N, FONTES H, et al. Wi-Fi rate adaptation using a simple deep reinforcement learning approach[C]// Proceedings of the 2022 IEEE Symposium on Computers and Communications (ISCC). Piscataway: IEEE Press, 2022: 1-3.

[25] CHEN S C, LI C Y, CHIU C H. An experience driven design for IEEE 802.11ac rate adaptation based on reinforcement learning[C]// Proceedings of the IEEE INFOCOM 2021-IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2021: 1-10.

[26] YE X W, FU L Q, CIOFFI J M. Joint codebook selection and MCS adaptation for mmWave eMBB services based on deep reinforcement learning[J]. IEEE Internet of Things Journal, 2024, 11(19): 31545-31560.

[27] GAO W, ZHENG P, HU Y L, et al. A novel link adaptation approach for URLLC: a DRL-based method with OLLA[C]//Proceedings of the 2024 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2024: 1-6.

[28] IEEE. Draft standard for information technology-telecommunications and information exchange between systems local and metropolitan area networks-specific requirements-part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 1: enhancements for high efficiency WLAN: IEEE P802.11ax/D5.0[S]. 2019.

[29] PORAT R.11ax evaluation methodology[S]. Doc: IEEE 802-11-14/0571r12, 2016.

[30] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q-learning[C]//Proceedings of 30th Association-for-the-Advancement-of-Artificial-Intelligence (AAAI) Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2016: 2094-2100.

[31] WANG Z Y, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning[C]//Proceedings of the 33rd International Conference on Machine Learning. International Machine Learning Society, 2016: 1995-2003.

[32] WATKINS C J C H. Learning from delayed rewards[J]. Robotics & Autonomous Systems, 1989.

[33] YIN H, LIU P, LIU K, et al. Ns3-Ai: fostering artificial intelligence algorithms for networking research[C]//Proceedings of the 2020 Workshop on ns-3. ns-3 Consortium, 2020: 57-64.

[34] ALBAR R, ARIF T Y, MUNADI R. Modified rate control for collision-aware in Minstrel-HT rate adaptation algorithm [C]// Proceedings of the 2018 International Conference on Electrical Engineering and Informatics (ICELTICs). Piscataway: IEEE Press, 2018:7-12.

[作者简介]

吴婷婷（2000– ），女，西南交通大学信息科学与技术学院硕士生，主要研究方向为Wi-Fi网络。

方旭明（1962– ），男，博士，西南交通大学信息科学与技术学院教授，主要研究方向为通信感知计算一体化网络、Wi-Fi网络、智能交通移动通信系统等。