基于深度强化学习的 OFDM 自适应导频设计

刘乔寿[1,2,3]，周雄[1,2,3]，刘爽[1,2,3]，邓义锋[1,2,3]

（1. 重庆邮电大学通信与信息工程学院，重庆 400065；2. 先进网络与智能互联技术重庆市高校重点实验室，重庆 400065；
3. 泛在感知与互联重庆市重点实验室，重庆 400065）

摘 要：针对正交频分复用系统，提出了一种基于深度强化学习的自适应导频设计算法。将导频设计问题映射为马尔可夫决策过程，导频位置的索引定义为动作，用基于减少均方误差的策略定义奖励函数，使用深度强化学习来更新导频位置。根据信道条件自适应地动态分配导频，从而利用信道特性对抗信道衰落。仿真结果表明，所提算法在 3GPP 的 3 种典型多径信道下相较于传统导频均匀分配方案信道估计性能有显著的提升。

关键词：正交频分复用；深度强化学习；马尔可夫决策过程；多径信道

中图分类号：TN92

文献标志码：A

DOI: 10.11959/j.issn.1000−436x.2023169

Adaptive pilot design for OFDM based on deep reinforcement learning

LIU Qiaoshou[1,2,3], ZHOU Xiong[1,2,3], LIU Shuang[1,2,3], DENG Yifeng[1,2,3]

1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
2. Advanced Network and Intelligent Connection Technology Key Laboratory of Chongqing Education Commission of China, Chongqing 400065, China
3. Chongqing Key Laboratory of Ubiquitous Sensing and Networking, Chongqing 400065, China

Abstract: For orthogonal frequency division multiplexing (OFDM) systems, an adaptive pilot design algorithm based on deep reinforcement learning was proposed. The pilot design problem was formulated as a Markov decision process, where the index of pilot positions was defined as actions. A reward function based on mean squared error (MSE) reduction strategy was formulated, and deep reinforcement learning was employed to update the pilot positions. The pilot was adaptively and dynamically allocated based on channel conditions, thereby utilizing channel characteristics to combat channel fading. The simulation results show that the proposed algorithm has significantly improved channel estimation performance compared with the traditional pilot uniform allocation scheme under three typical multipath channels of 3GPP.

Keywords: OFDM, deep reinforcement learning, Markov decision process, multipath channel

0 引言

正交频分复用（OFDM, orthogonal frequency division multiplexing）技术目前已广泛应用于 4G 和 5G，并将作为关键技术应用于 6G。在无线通信系统中，发送端发射的信号通常会因为无线信道的特性而失真。为了在接收端恢复发射信号，解调前必须对信道特性进行补偿。在不断变化的无线信道传输条件下，能够快速且准确地估计出信道状态信息（CSI, channel state information）是保证通信系统高吞吐量的关键。信道估计是获取 CSI 的重点，其中基于导频辅助的信道估计算法由于可以与 OFDM

收稿日期：2023−05−26；**修回日期：**2023−08−10

通信作者：周雄，zhouxiongz@163.com

基金项目：国家自然科学基金资助项目（No.61901075）；重庆市教委科学技术基金资助项目（No.KJZDK202200604）

Foundation Items: The National Natural Science Foundation of China (No.61901075), The Science and Technology Research Program of Chongqing Municipal Education Commission (No.KJZDK202200604)

系统很好地结合得到了广泛的关注和应用。发射端在帧内某些特定符号的特定子载波上发送已知的导频信息，接收端收到这些导频信息后计算出导频子载波处的 CSI，再通过常用的信道估计算法，如最小二乘（LS, least square）和线性最小均方误差（LMMSE, linear minimum mean square error）等，估计出数据部分的 CSI。

信道估计的性能与导频数量和导频图案的设计密切相关[1-4]。传统的无线通信系统在进行信道估计时一般使用均匀分配的导频图案，常用的均匀导频图案有块状导频、梳状导频等。均匀导频分配方案在面对复杂多变的多径信道时由于导频间隔固定，因此通常不是最优的导频分配方案，这导致其不能很好地对抗信道衰落，性能表现一般。针对上述问题，文献[1]提出一种基于深度学习的导频图案设计和信道估计联合算法，使用特征选择方法 ConcreteAE 寻找数据传输中包含信息最多的位置。文献[2]提出一种有效的导频剪枝技术，通过在训练过程中从密集的神经网络层中修剪不重要的神经元降低系统开销。文献[3]研究了 OFDM 系统的导频设计和信号检测问题。文献[4]针对短帧内导频数量有限的信道估计问题，对帧结构的块长度和导频长度进行了联合优化。从以上文献可知，即使已有文献进行导频设计方面的研究，其研究结果可以在一定程度上克服固定导频间隔引发的问题，但是依然存在不足。例如，基于深度学习的方法设计非均匀导频方案，该方案虽然摆脱了导频均匀分配带来的问题，但是需要大量可靠的训练数据集。同时，许多基于人工智能的导频设计方案虽然相比传统导频均匀分配方案的性能有所提升，但是它并没有脱离固定导频间隔的范畴，这类方案不一定是最优的导频分配模式。目前已有的研究并不能解决实际通信系统中信道动态变化导致固定导频设计信道估计性能下降的问题，因此寻求一种根据信道动态变化自适应调整导频图案的方法至关重要。

目前，自适应导频间隔[5-6]可以有效解决非平稳信道对信道估计带来的问题。文献[5]提出了一种基于深度 Q 网络（DQN, deep Q-network）的学习算法，对导频间隔和导频功率同时进行优化，在降低系统成本的同时获得最大的信道估计性能。文献[6]设计了一种基于码本的方法调整导频间隔和导频功率。但是基于自适应导频间隔的方案依然属于固定导频间隔的范畴，不一定是最优的导频分配模式。

由于深度强化学习在动态频谱接入[7-9]、任务卸载[10-12]和网络安全[13-14]等通信领域发挥出的强大实力，已经有相关学者将其应用到信道估计[15-16]和导频序列选择[17]方面的研究。文献[15]提出一种基于 Q 学习的信道估计去噪算法，在 LS 估计的基础上进行去噪，将去噪机制定义为一个马尔可夫决策过程，获得接近 LMMSE 估计的性能。文献[16]提出一种基于深度确定性策略梯度（DDPG, deep deterministic policy gradient）的端到端信道预测和波束成形算法，采用行动者−评论家网络，在没有完美 CSI 的前提下，其信道预测能力优于最小均方误差（MMSE, minimum mean square error）信道估计算法。文献[17]针对非授权多址（GFMA, grant-free multiple access）系统提出一种基于深度强化学习的导频序列选择方案，以此减轻不同用户的导频序列冲突。考虑深度强化学习的最优策略寻找能力以及现有导频设计方法存在的缺陷，本文将深度强化学习应用到 OFDM 系统导频设计中。

本文提出了一种基于深度强化学习的自适应导频设计算法，将导频设计问题映射为马尔可夫决策过程，将导频位置的索引定义为动作，用基于减少均方误差（MSE, mean square error）的策略定义奖励函数，使用深度强化学习算法更新导频位置。所提出的自适应导频设计算法可以利用信道的特性找到对其来说最优的导频分配模式，以此更好地对抗信道衰落，得到更佳的信道估计性能。值得一提的是，本文就所提自适应导频设计算法在动态信道下的性能也展开了研究。本文的主要贡献如下。

1) 使用深度强化学习求解导频设计问题。为了寻求 OFDM 系统下信道估计性能最优的导频分配方案，本文将 OFDM 系统中导频设计问题建模为马尔可夫决策过程，采用深度强化学习求解该模型得到导频分配方案。

2) 提出的基于深度强化学习的自适应导频设计算法根据信道条件自适应地动态分配导频。通过仿真测试所提算法在动态信道下的收敛性能，证明所提算法可以根据信道条件自适应地分配导频，以此利用信道特性对抗信道衰落。

3) 验证了所提算法的有效性。通过仿真分析所提算法与对比算法在 OFDM 系统中的信道估计性能，证明了所提算法在 OFDM 系统中具有优秀的信道估计性能。

1 系统模型

本文考虑一个具有 N 个子载波的 OFDM 系统，子载波索引 $n \in \{1,2,\cdots,N\}$，如图 1 所示。

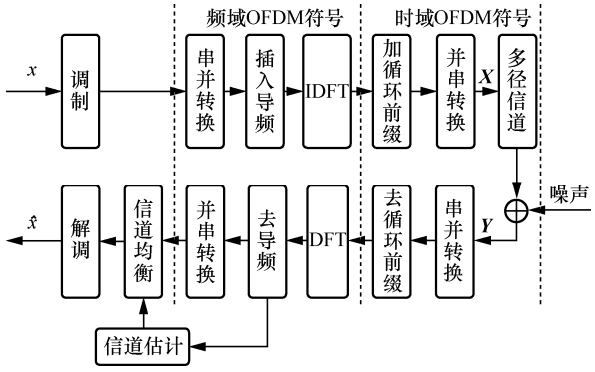


图 1 OFDM 系统

二进制比特流 x 经过调制、串并转换、插入导频、离散傅里叶反变换（IDFT, inverse discrete Fourier transform）、加循环前缀、并串转换、多径信道后到达接收端；接收端对接收到的信号进行串并转换、去循环前缀、离散傅里叶变换（DFT, discrete Fourier transform）、去导频（同时进行信道估计）、并串转换、信道均衡、解调后得到二进制比特流 $\hat{x}$。图 1 中发射端和接收端数据的关系如式(1)所示。

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{W} \tag{1}$$

其中，$\boldsymbol{Y} = [Y_1, Y_2, \cdots, Y_N]$ 表示接收端频域 OFDM 符号，$\boldsymbol{X} = [X_1, X_2, \cdots, X_N]$ 表示发射端频域 OFDM 符号，$\boldsymbol{H} = [H_1, H_2, \cdots, H_N]$ 表示实际的信道矩阵，$\boldsymbol{W} = [W_1, W_2, \cdots, W_N]$ 表示噪声矩阵，$W_n \sim \mathrm{CN}(0, \sigma_w^2)$ 表示均值为 0、方差为 σ_w^2 的加性白高斯噪声。

OFDM 系统采用基于导频辅助的信道估计方案。发射端发送包含已知导频序列的数据给接收端，接收端利用导频序列计算导频处的 CSI，然后根据信道估计算法估计出数据部分的 CSI，得到的估计信道矩阵为 $\hat{\boldsymbol{H}} = [\hat{H}_1, \hat{H}_2, \cdots, \hat{H}_N]$。信道估计算法考虑了 LS 和 LMMSE 算法。LS 算法是一种较简单的信道估计算法，没有考虑噪声的影响，准确性相对有限。它的目标是使信道估计值与实际值的均方误差最小，如式(2)所示。

$$\hat{\boldsymbol{H}}_{\mathrm{LS}} = \boldsymbol{X}^{-1}\boldsymbol{Y} \tag{2}$$

LMMSE 算法是在 MMSE 算法的基础上改进而来的，如式(3)所示。

$$\hat{\boldsymbol{H}}_{\mathrm{LMMSE}} = \boldsymbol{R}_{\mathrm{HH}}\left[\boldsymbol{R}_{\mathrm{HH}} + \frac{\beta}{\mathrm{SNR}}\boldsymbol{I}_{N\times N}\right]^{-1}\hat{\boldsymbol{H}}_{\mathrm{LS}} \tag{3}$$

其中，调制因子 β 与选用的调制方式有关，$\boldsymbol{R}_{\mathrm{HH}}$ 是信道的自相关矩阵，SNR 是信噪比，$\boldsymbol{I}_{N\times N}$ 是 $N \times N$ 的单位矩阵。LMMSE 估计在最小化 MSE 方面有着较好的性能，但是它的计算复杂度比较高，并且需要信道统计信息的先验知识。

传统导频均匀分配方案主要分为梳状、块状以及混合状。梳状导频在频域上等间隔地插入导频序列，在频域内进行信道估计，在接收端对收到的导频符号进行频域插值运算。梳状导频在时域上是连续的，可以有效对抗时间选择性衰落，对频率选择性衰落信道比较敏感。块状导频是指在一个 OFDM 符号内的所有子载波上都插入导频，导频在频域内是连续的，在时域上是等间隔的，它可以有效对抗频率选择性衰落，对时间选择性衰落信道比较敏感。混合状导频是梳状导频和块状导频结合之后的折中，在时域和频域上相隔一定间隔插入导频符号，导频符号在时域和频域的分布都是离散的。

OFDM 系统计算导频位置的 CSI 后，进行信道插值计算数据部分的 CSI。导频位置对信道插值结果至关重要，导频位置不当会丢失重要节点的数据造成信道估计准确度降低。然而，导频均匀分配方案没有考虑导频位置对信道插值的影响，不能有效地结合信道特性分配导频，往往不是当前信道模型下的最优导频分配模式。因此，本文在梳状导频的基础上，结合深度强化学习在寻找最优策略上的优势，提出了一种自适应导频设计算法，考虑了一个基于帧的传输场景，一帧数据的完整传输时间为一个时隙，其中每个帧由 M 个 OFDM 符号组成。该系统旨在从导频序列中估计信道，以正确检测同一帧内的数据符号。

2 算法设计

2.1 马尔可夫决策过程模型

一般来说，基于强化学习的问题求解被认为是智能体根据环境状态在一系列时隙里依次选择动作进行学习的过程。强化学习算法是基于马尔可夫决策过程公式发展来的，该公式包括状态空间 $\boldsymbol{S}$、动作空间 $\boldsymbol{A}$，即时奖励函数 $\boldsymbol{R}: \boldsymbol{S} \times \boldsymbol{A} \to \mathbb{R}$ 和状态转移概率集合 $\boldsymbol{P}(\boldsymbol{S}, \boldsymbol{A})$。在所提自适应导频设计算法中，状态空间 $\boldsymbol{S}$ 是一帧数据所需要二进制比特流的集合，动作空间 $\boldsymbol{A}$ 是所有可能导频位置索引的集

合，即时奖励函数如式(4)所示。

$$r_t = -\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(\left|\hat{H}_{mnt} - H_{mnt}\right|^2\right) \tag{4}$$

其中，$\hat{H}_{mnt}$ 是第 t 帧数据中第 m 个 OFDM 符号第 n 个子载波的估计 CSI，H_{mnt} 是第 t 帧数据中第 m 个 OFDM 符号第 n 个子载波的实际 CSI。式(4)中的负号用于保证估计误差越小，奖励越高。对于本文提出的自适应导频设计算法，智能体在每个时隙都遵循环境的策略与环境进行交互，优化目标即总回报，如式(5)所示。

$$G_t = \sum_{l=t}^{\infty}\gamma^{l-t}r_{l+1} = -\sum_{l=t}^{\infty}\gamma^{l-t}\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(\left|\hat{H}_{mn(l+1)} - H_{mn(l+1)}\right|^2\right) \tag{5}$$

其中，γ 是折现因子，$\gamma \in [0,1]$。

强化学习智能体的目标是找到一个策略 π 来最大化总回报，π 表示一个从状态空间 $\boldsymbol{S}$ 到动作空间 $\boldsymbol{A}$ 的函数映射。马尔可夫决策过程模型的求解依赖 2 个函数，即状态价值函数 $V_\pi(s) = \mathrm{E}_\pi[G_t \mid S_t = s]$ 和动作价值函数 $Q_\pi(s,a) = \mathrm{E}_\pi[G_t \mid S_t = s, A_t = a]$，其中 $E_\pi[\cdot]$ 表示遵循策略 π 的期望。智能体最大化总回报的目标等价于寻找一个最优策略 π^* 使任意状态下的 $V_\pi(s)$ 最大化，其选择的最优动作是 $\max\limits_a Q_{\pi^*}(s,a)$，最优策略 π^* 的动作价值函数为 $Q_{\pi^*}(s,a)$。所提自适应导频设计算法的目的是获得最优导频分配策略 π^*，得到最优的导频分配模式，其优化目标是最大化总回报 $J_1(\pi) = \mathrm{E}_\pi[G_t]$。本文求解导频设计问题可以表述为

$$\max_{\pi,\hat{\boldsymbol{H}}} J_1(\pi) = \mathrm{E}_\pi[G_t] \tag{6}$$

2.2　基于 Dueling DQN 的自适应导频设计

针对离散的动作空间一般使用基于价值的强化学习算法，连续的动作空间则一般使用基于策略的强化学习算法。本文提出的自适应导频设计算法中采用的动作空间是离散的，因此使用的深度强化学习算法是 DQN 的变体算法 Dueling DQN[18]。

Dueling DQN 在 DQN 的基础上改进网络结构，更详细地描述动作价值函数，解决了高估动作价值的问题。基于 Dueling DQN 自适应导频设计算法的网络结构如图 2 所示。

将 OFDM 系统视为环境，Dueling DQN 与其进行信息交互。Dueling DQN 由特征层、状态价值层和优势层三部分组成。特征层由三层全连接层 F_1、F_2 和 F_3 组成。状态价值层和优势层分别由两层全连接层 S_1、S_2 和 A_1、A_2 组成。Dueling DQN 的网络参数为 θ。Dueling DQN 在同一状态计算不同动作

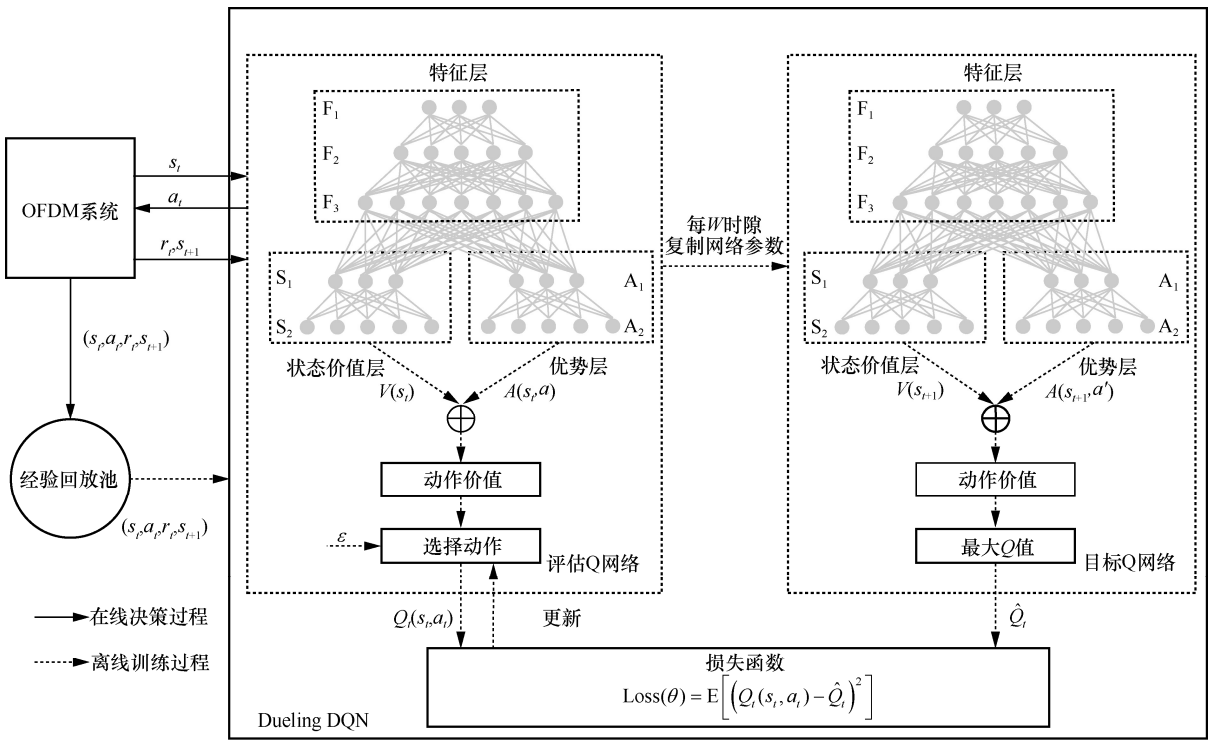


图 2　基于 Dueling DQN 自适应导频设计算法的网络结构

的优势函数时不重复计算状态价值，提高了动作价值的更新效率。在时隙 t，Dueling DQN 观察到当前环境状态 s_t 后，利用特征层提取状态 s_t 的特征。Dueling DQN 的特征层后有 2 个分支，即状态价值层和优势层，两者会根据状态 s_t 分别计算状态价值 $V(s_t)$ 和动作优势函数值 $A(s_t,a)$，再将 2 个值结合在一起计算所有可用动作的动作价值，计算方法如下[15]

$$Q_t(s_t,a)=V(s_t)+\left(A(s_t,a)-E[A(s_t,a)]\right) \tag{7}$$

Dueling DQN 采取了值为 ε 的贪心策略，每次选择动作时生成一个 0～1 的随机数，如果该数小于 ε，使用式(8)选择动作 a_t，反之则从动作空间随机选择动作 a_t。

$$a_t=\arg\max_a Q_t(s_t,a) \tag{8}$$

OFDM 系统根据 a_t 得到插入导频的位置，接收端根据导频进行信道估计。Dueling DQN 收到奖励 r_t 并观测到下一状态 s_{t+1}，目标动作价值 $\hat{Q}_t$ 为

$$\hat{Q}_t=r_t+\gamma\max_{a'}Q_t(s_{t+1},a') \tag{9}$$

Dueling DQN 的损失函数如式(10)所示，最小化损失函数对模型进行训练，并通过梯度下降法逐步更新所有网络参数，表示为

$$\mathrm{Loss}(\theta)=\mathrm{E}\left[\left(Q_t(s_t,a_t)-\hat{Q}_t\right)^2\right] \tag{10}$$

$$\theta\leftarrow\theta-\eta\nabla\theta \tag{11}$$

其中，η 为学习率。

Dueling DQN 采用了一个容量为 C 的经验回放池，在与环境的信息交互中，将 $e_t=(s_t,a_t,r_t,s_{t+1})$ 的元组数据存入经验回放池。待经验回放池存满元组数据后，为了防止数据相关性导致网络的假收敛会打乱其中元组数据的顺序，再从中抽取数据进行训练。Dueling DQN 有 2 个网络结构一样的网络，分别为评估 Q 网络和目标 Q 网络，其输出分别是 $Q_t(s_t,a_t)$ 和 $\hat{Q}_t$，因此式(10)中的 2 个参数是不断变化的，会导致 Dueling DQN 难以收敛。为了防止该问题的发生，Dueling DQN 会固定目标 Q 网络的参数，待迭代 W 步后才更新目标 Q 网络的参数。本文将基于 Dueling DQN 的自适应导频设计算法总结在算法 1 中。

算法 1 基于 Dueling DQN 的自适应导频设计算法

初始化 经验回放池的容量 C、总学习回合 N_{epochs}、一个学习回合总帧数 T、评估 Q 网络参数 θ 和目标 Q 网络参数 θ^- 和目标 Q 网络更新步长 W

1) for 学习回合 epochs $=1,2,\cdots,N_{\text{epochs}}$
2) 初始化获取状态 s_0
3) for frames $t=1,2,\cdots,T$ do
4) OFDM 系统生成一帧二进制比特流 s_t
5) 以 ε 的概率选择 $a_t=\arg\max_a Q_t(s_t,a)$，否则随机选择一个动作 a_t
6) 执行动作 a_t 得到奖励 r_t 和下一个状态 s_{t+1}
7) 存储 (s_t,a_t,r_t,s_{t+1}) 到经验回放池 D 中，$s_t\leftarrow s_{t+1}$
8) if 经验回放池的当前容量 $>C$
9) 从经验回放池中随机采样 batch-size 样本 (s,a,r,s')
10) 根据式(9)计算目标动作价值 $\hat{Q}_t$
11) 最小化网络损失函数式(10)
12) 根据式(11)执行梯度下降，更新评估 Q 网络参数 θ
13) 每 W 时隙更新目标 Q 网络参数 θ^-
14) end if
15) end for
16) end for

此外，本文对动作空间 $\boldsymbol{A}$ 进行剪枝操作，对所有可能的导频位置设置了最小导频间隔 $d_{\min}$ 和最大导频间隔 $d_{\max}$。OFDM 系统相邻子载波的 CSI 是比较接近的，$d_{\min}$ 过小对信道估计性能几乎没有提升还极大地浪费导频资源。合适的 $d_{\min}$ 不仅避免浪费导频资源，还会对 Dueling DQN 进行良好的初始化。设置 $d_{\max}$ 让导频在合适范围里寻找 CSI 波动比较大的地方，同时不会干扰其他导频的位置选择，减轻 Dueling DQN 负担，使 Dueling DQN 能够更快收敛。因此，$d_{\min}$ 和 $d_{\max}$ 是应该考虑的。

2.3 计算复杂度分析

深度强化学习的计算复杂度与选取的状态空间、动作空间以及网络结构相关。因此本文通过所提算法的网络参数来衡量计算复杂度。对于一个两层全连接神经网络来说，假设训练一个网络参数计算复杂度为 L，输入层和输出层神经元个数分别为 s_i 和 s_{i+1}，那么该两层全连接神经网络的计算复杂度为 $O\left((s_is_{i+1}+s_{i+1})L\right)$。Dueling DQN 中 F_1～F_3 层的神经元个数分别为 f_1～f_3，S_1 和 S_2 层的神经元个数分别为 c_1 和 c_2，A_1 和 A_2 层的神经元个数分别为 v_1 和 v_2，因此本文所提算法的计算复杂度为

$O\left((f_1f_2+f_2f_3+f_3c_1+f_3v_1+c_1c_2+v_1v_2+f_2+f_3+c_1+c_2+v_1+v_2)L\right)$。

3　仿真分析

本节通过仿真来说明所提算法的性能。在仿真中，OFDM 系统子载波个数 $N=64$，采样率为 0.96 Msymbol/s，子载波间隔为 15 kHz，一帧中 OFDM 符号数量 $M=7$，循环前缀长度为 16，导频数量 $P=4$，调制方式选用 QPSK。传统导频均匀分配方案采用固定导频间隔为 16 的梳状导频，因此在自适应导频设计算法中，导频间隔均值设置为 16，同时考虑到 Dueling DQN 初始化以及网络负担等因素，设置最小导频间隔 $d_{\min}=9$，最大导频间隔 $d_{\max}=23$。Dueling DQN 参数设置如表 1 所示，其中，网络层参数依次为网络类型（Linear 表示全连接层）、神经元个数和激活函数。根据 OFDM 子载波个数和调制阶数，Dueling DQN 中隐藏层神经元个数设置如下：f_2 为 64，f_3、c_1 和 v_1 为 128。这样设置能最大限度地将信道特征映射到 OFDM 符号上，以此从中选择最合适的导频位置。W 的取值关系到 Dueling DQN 能否正常收敛，W 取值过大，Dueling DQN 收敛速度较慢；W 取值过小，Dueling DQN 容易局部最优。折现因子 γ 设置为 1，原因是每一帧数据都同等重要，本文希望能够最小化数据传输中每一帧估计信道与实际信道的 MSE。在 Dueling DQN 训练中，设置总学习回合 $N_{\text{epochs}}=50$，一个学习回合总帧数 $T=10\,000$。

表 1　　Dueling DQN 参数设置

参数		值
网络层	F_1	Linear,840
	F_2	Linear,64,ReLU
	F_3	Linear,128,ReLU
	S_1	Linear,128,ReLU
	S_2	Linear,1
	A_1	Linear,128,ReLU
	A_2	Linear,54070
batch-size		256
经验回放池 C		5 000
W		5 000
ε		0.9
γ		1
η		0.01
优化器		Adam

本文测试了所提基于 Dueling DQN 的自适应导频设计算法在多径信道下的收敛性能。其中信道为 3GPP 协议中选取的 3 种频率选择性衰落信道 EPA、EVA 和 ETU[19]，信道参数如表 2 所示。

表 2　　信道参数

信道	时延/ns	功率增益/dB
EPA	0	0.0
	30	−1.0
	70	−2.0
	90	−3.0
	110	−8.0
	190	−17.2
	410	−20.8
	—	—
	—	—
EVA	0	0.0
	30	−1.5
	150	−1.4
	310	−3.6
	370	−0.6
	710	−9.1
	1 090	−7.0
	1 730	−12.0
	2 510	−16.9
ETU	0	−1.0
	50	−1.0
	120	−1.0
	200	0.0
	230	0.0
	500	0.0
	1 600	−3.0
	2 300	−5.0
	5 000	−7.0
	0	−1.0

将所提算法与传统导频均匀分配方案在 3 种多径信道下进行对比，测试两者在不同学习回合下的总回报，其中信噪比为 5 dB，选用 LS 估计。Adaptive 代表所提算法，Orginal 代表传统导频均匀分配方案。

所提算法和传统导频均匀分配方案在 EPA 信道下不同学习回合的总回报如图 3 所示。从图 3 中可以看出，所提算法相比传统导频均匀分配方案具有压倒性优势，其总回报结果一直领先于传统导频均匀分配方案。所提算法在第 2 个学习回合性能收

敛，总回报为−3 925，传统导频均匀分配方案总回报为−5 311，所提算法比传统导频均匀分配方案总回报提升了 26%，证明了基于 Dueling DQN 的自适应导频设计算法的可行性。

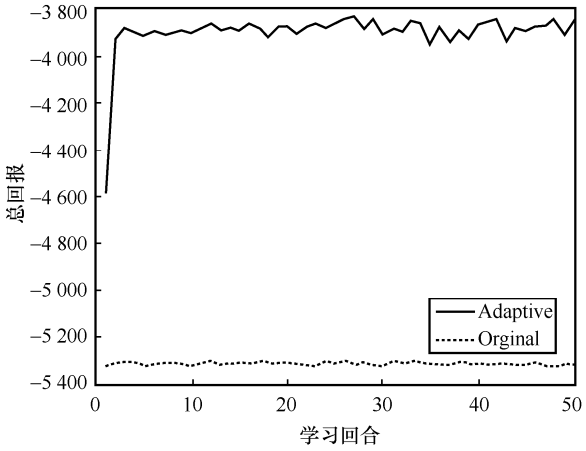


图 3 EPA 信道下不同学习回合的总回报

所提算法和传统导频均匀分配方案在 EVA 信道下不同学习回合的总回报如图 4 所示。从图 4 中可以看出，所提算法依然全程领先传统导频均匀分配方案。所提算法在第 5 个学习回合时总回报收敛，总回报的数值为−17 644，而传统导频均匀分配方案总回报为−27 207，所提算法在 EVA 信道下比传统导频分配方案总回报提升了 35%。

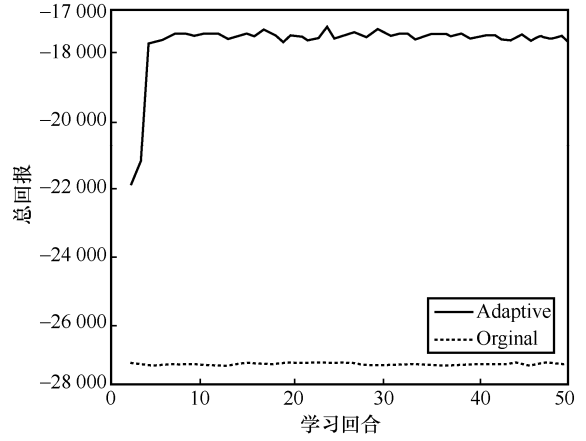


图 4 EVA 信道下不同学习回合的总回报

所提算法和传统导频均匀分配方案在 ETU 信道下不同学习回合的总回报如图 5 所示。从图 5 可以看出，所提算法在第 7 个学习回合时总回报收敛，数值为−34 135，而传统导频均匀分配方案的总回报为−46 283，所提算法在 ETU 信道下比传统导频分配方案总回报提升了 26%。

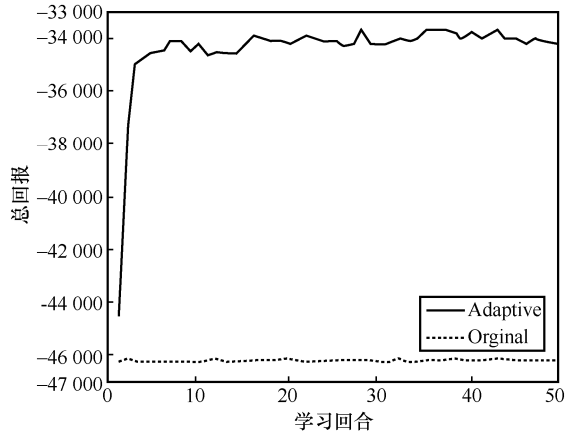


图 5 ETU 信道下不同学习回合的总回报

从图 3～图 5 可以看出，所提算法在 EPA 信道下收敛最快，EVA 信道次之，ETU 信道最慢，这是因为不同信道的信道特征复杂程度不同，EPA 信道的信道特征复杂程度最低，EVA 次之，ETU 复杂程度最高，所提算法需要花费不同的时间去学习它们的信道特征。所提算法在不同信道中都能收敛，并得到远优于传统导频均匀分配方案的总回报，证明了所提算法的有效性。

本文进一步测试了动态信道下的收敛性能，所提算法和传统导频均匀分配方案在动态信道下不同学习回合的总回报如图 6 所示。将所提算法与传统导频均匀分配方案在动态信道下进行对比，信噪比为 5 dB，采用 LS 估计。动态信道开始为 EPA 信道，在第 21 个学习回合开始时变为 EVA 信道，在第 41 个学习回合开始时变为 ETU 信道。从图 6 可以看出，所提算法和传统导频均匀分配方案在第 21 个学习回合由于信道改变，得到的总回报大幅度下降。所提算法在第 21 个学习回合得到的总回报远大于传统导频均匀分配方案，且在第 22 个学习回合就达到收敛。同样地，在第 41 个学习回合，信道再次改变，所提算法只需要 2 个学习回合就能在 ETU 信道下收敛。所提出的基于 Dueling DQN 的自适应导频设计算法面对信道的变化依然能表现出优秀的收敛性能，证明了该算法的自适应性。

为了更加直观地说明所提算法的收敛性能，本文测试了所提算法和传统导频均匀分配方案在动态信道下的 MSE，如图 7 所示。将所提算法与传统导频均匀分配方案在动态信道下进行对比，信噪比为 5 dB，采用 LS 估计，横坐标表示时间，纵坐标表示实际 CSI 与估计 CSI 的 MSE。动态信道开始为 EPA 信道，在第 100 001 个时隙变为 EVA 信道，在

第 200 001 个时隙变为 ETU 信道。从图 7 可以看出，所提算法的 MSE 在动态信道下可以很好地收敛，且收敛之后的数值远小于传统导频均匀分配方案，进一步验证了所提算法的有效性。图 7 中 EPA 信道下前 5 000 帧，所提算法的 MSE 在传统导频均匀分配方案的 MSE 附近波动，这是因为所提算法在训练早期与环境交互对动作空间进行探索，会不断试错来找到最优策略，此时所提算法还未收敛。等到所提算法逐渐收敛，会展现远低于传统导频均匀分配方案的 MSE。但是，所提算法依然存在缺陷，需要一定的训练时间和计算资源才能发挥良好的信道估计效果，在快速变化的动态信道下展现的信道估计性能会降低。

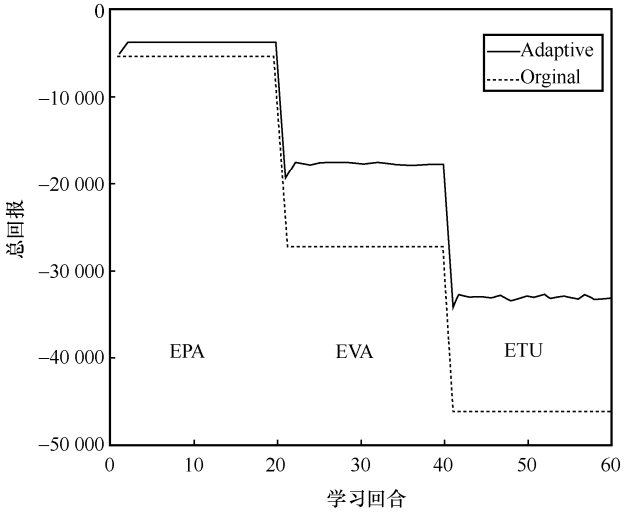


图 6　所提算法和传统导频均匀分配方案在动态信道下不同学习回合的总回报

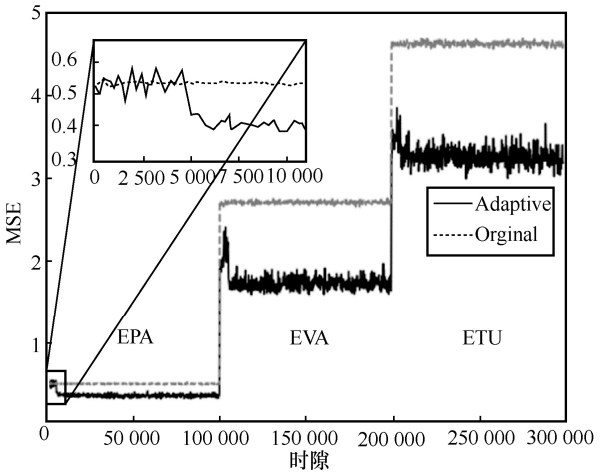
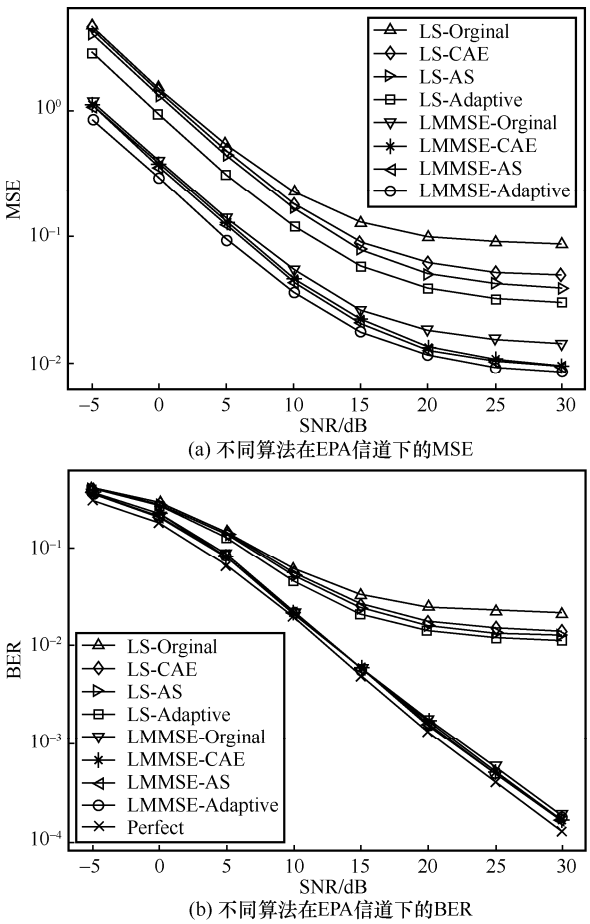


图 7　所提算法和传统导频均匀分配方案在动态信道下的 MSE

本文在 EPA、EVA 和 ETU 这 3 种信道下对比了 OFDM 系统分别采用所提自适应导频设计算法、基于 ConcreteAE 的导频设计算法[1]、基于文献[5]的自适应导频间隔算法和传统导频均匀分配方案的性能，评价指标是估计信道和实际信道的 MSE 和 OFDM 系统的误码率（BER, bit error rate），信噪比范围为−5～30 dB，采用 LS 估计和 LMMSE 估计。LS-Orginal 表示采用传统导频均匀分配方案和 LS 估计，LS-CAE 表示采用基于 ConcreteAE 的导频设计算法和 LS 估计，LS-AS 表示采用基于文献[5]的自适应导频间隔算法和 LS 估计，LS-Adaptive 表示采用所提自适应导频设计算法和 LS 估计，LMMSE-Orginal 表示采用传统导频均匀分配方案和 LMMSE 估计，LMMSE-CAE 表示采用基于 ConcreteAE 的导频设计和 LMMSE 估计，LMMSE-AS 表示采用基于文献[5]的自适应导频间隔算法和 LMMSE 估计，LMMSE-Adaptive 表示采用所提自适应导频设计算法和 LMMSE 估计，Perfect 表示采用完美的信道估计。不同算法在 EPA 信道下的 MSE 和 BER 如图 8 所示。



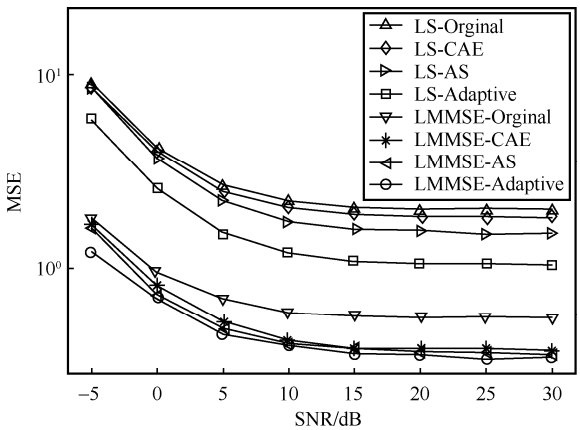
(a) 不同算法在EPA信道下的MSE

(b) 不同算法在EPA信道下的BER

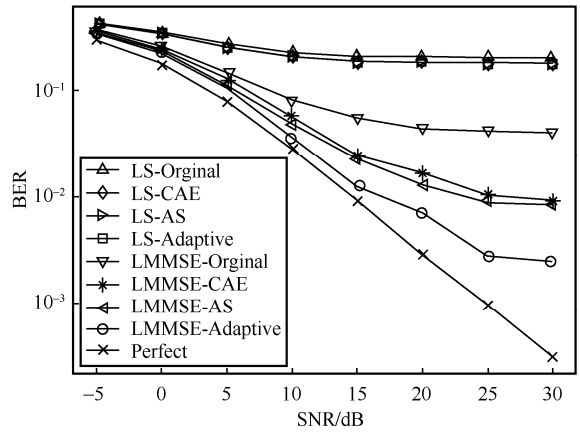
图 8　不同算法在 EPA 信道下的 MSE 和 BER

从图 8(a)可以看出，在相同的信道估计算法下，所提自适应导频设计表现的 MSE 性能最好，基于文献[5]的自适应导频间隔算法和基于 ConcreteAE 的导频设计算法次之，传统导频均匀分配方案最差。当信噪比为 5 dB 时，LS-Adaptive 的 MSE 为 0.31，LS-AS 的 MSE 为 0.44，LS-CAE 的 MSE 为 0.47，LS-Orginal 的 MSE 为 0.53。此时 LS-Adaptive 相较于 LS-AS 有 30%的 MSE 性能提升，相较于 LS-CAE 有 34%的 MSE 性能提升，相较于 LS-Orginal 有 42%的 MSE 性能提升，证明了所提算法在 MSE 性能上有明显的优势。所提算法相较于基于文献[5]的自适应导频间隔算法有明显的 MSE 性能提升，这是因为所提算法具有更大的动作空间，有更大可能包括了信道估计性能最佳的导频方案，但是这样带来的缺点是所提算法需要更长的训练时间和更多的计算资源。从图 8(b)可以看出，当信道估计算法为 LS 估计时，本文所提出的自适应导频设计相较于另外 3 种导频方案有明显的性能提升。而在 LMMSE 估计时，自适应导频设计相较于另外 3 种导频方案只有微弱的优势。这是因为 LMMSE 有信道的先验信息，性能比 LS 估计好，且 EPA 信道的复杂程度较低，即使采用传统导频均匀分配方案也有很好的 BER 性能。而 LS 估计没有信道的先验信息，更加依赖于导频位置的选取。由于所提算法需要一定的训练时间和计算资源，并不适于导频位置对信道估计性能没有明显影响的情况。对于导频位置对信道估计性能有明显影响的情况，使用所提算法牺牲一定计算资源可以有效提升信道估计性能。

不同算法在 EVA 信道下的 MSE 和 BER 如图 9 所示。从图 9(a)可以看出，在相同的信道估计算法下，所提自适应导频设计算法的 MSE 性能全程领先于基于文献[5]的自适应导频间隔算法、基于 ConcreteAE 的导频设计算法和传统导频均匀分配方案。当信噪比为 10 dB 时，LS-Adaptive 的 MSE 为 1.19，LS-AS 的 MSE 为 1.73，LS-CAE 的 MSE 为 2.05，LS-Orginal 的 MSE 为 2.23。当信噪比为 10 dB 时，LS-Adaptive 相较于 LS-AS 有 31%的 MSE 性能提升，相较于 LS-CAE 有 42%的 MSE 性能提升，相较于 LS-Orginal 有 47%的 MSE 性能提升。在信噪比为 15 dB 时，几种方法的 MSE 性能都趋于平缓。从图 9(b)可以看出，在相同的信道估计算法下，所提自适应导频设计算法的 BER 性能全程领先于基于文献[5]的自适应导频间隔算法、基于 ConcreteAE 的导频设计算法和传统导频均匀分配方案。当信噪比为 15 dB 时，LMMSE- Adaptive 的 BER 为 0.012，LMMSE-AS 的 BER 为 0.022，LMMSE-CAE 的 BER 为 0.024，LMMSE-Orginal 的 BER 为 0.054。LMMSE-Adaptive 相较于 LMMSE-AS 有 45%的性能提升，相较于 LMMSE-CAE 有 50%的性能提升，相较于 LMMSE-Orginal 有 78%的性能提升。在 LS 信道估计下，所提自适应导频设计算法相较于另外 3 种导频方案只能展现微弱的优势。这是因为 EVA 信道复杂程度较高，LS 估计算法不能展现良好的 BER 性能，此时导频位置对于 BER 性能的影响比较低。而在 LMMSE 估计下，导频位置的选取对于 BER 性能有着明显的影响。因此，所提算法适用于导频位置的选取对信道估计性能有明显影响的情况。在该种情况下，所提算法可以展现优秀的信道估计性能。



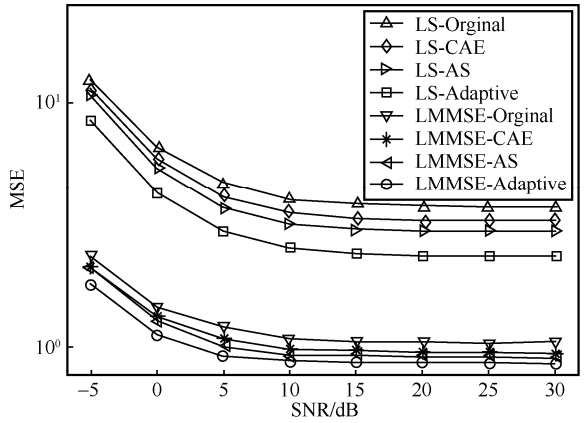
(a) 不同算法在EVA信道下的MSE



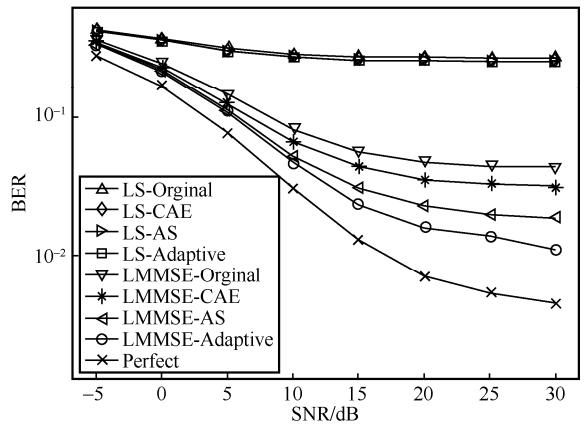
(b) 不同算法在EVA信道下的BER

图 9 不同算法在 EVA 信道下的 MSE 和 BER

不同算法在 ETU 信道下的 MSE 和 BER 如图 10 所示。从图 10 可以看出，与在 EPA 信道和 EVA 信道一样，在相同的信道估计算法下，所提自适应导频设计算法的 MSE 和 BER 性能都优于对比算法。



(a) 不同算法在ETU信道下的MSE



(b) 不同算法在ETU信道下的BER

图 10　不同算法在 ETU 信道下的 MSE 和 BER

通过观察图 8～图 10 可以发现，所提自适应导频设计算法在 EPA、EVA 和 ETU 信道下都能展现优秀的 MSE 和 BER 性能，相较于基于文献[5]的自适应导频间隔算法、基于 ConcreteAE 的导频设计算法和传统导频均匀分配方案有着显著优势。

最后，本文对比了所提算法和另外 3 种算法的计算复杂度。本文通过神经网络参数衡量计算复杂度，所提算法的计算复杂度为 $O(7\ 070\ 327L)$，基于文献[5]的自适应导频间隔算法的计算复杂度为 $O(353\ 935L)$，基于 ConcreteAE 的导频设计算法的计算复杂度为 $O(349\ 694L)$。由于传统导频均匀分配方案没有涉及神经网络，因此传统导频均匀分配方案的计算复杂度为常数级别 $O(1)$。可以看到，本文提出的计算复杂度最高，是基于 ConcreteAE 的导频设计算法计算复杂度的 20 倍。这是因为所提算法具有较大的动作空间，需要一定的训练时间和计算资源才能寻找到最优的导频分配方案。因此，所提算法并不适用于计算资源匮乏的通信系统。但对于具有充足计算资源且对信道估计性能要求高的通信系统来说，所提算法是一个比较好的选择。

4　结束语

本文针对 OFDM 系统的导频设计问题展开研究，提出一种基于深度强化学习的自适应导频设计算法。所提自适应导频设计算法利用深度强化学习的最优策略寻找能力，根据不同的信道条件自适应地得到不同的导频分配方案，以此更有效地对抗信道衰落。最后，通过仿真分析所提出的自适应导频设计算法在 EPA、EVA 和 ETU 这 3 种信道以及动态信道下的收敛性能，证明了其可行性；对比所提算法和另外 3 种算法在 3 种多径信道下的 MSE 和 BER，证明了所提算法优秀的信道估计性能。

参考文献：

[1] SOLTANI M, POURAHMADI V, SHEIKHZADEH H. Pilot pattern design for deep learning-based channel estimation in OFDM systems[J]. IEEE Wireless Communications Letters, 2020, 9(12): 2173-2176.

[2] MASHHADI M B, GÜNDÜZ D. Pruning the pilots: deep learning-based pilot design and channel estimation for MIMO-OFDM systems[J]. IEEE Transactions on Wireless Communications, 2021, 20(10): 6315-6328.

[3] CHEN H, ZHANG Q Q, LONG R Z, et al. Pilot design and signal detection for symbiotic radio over OFDM carriers[C]//Proceedings of IEEE Global Communications Conference. Piscataway: IEEE Press, 2023: 1887-1892.

[4] CAO J, ZHU X, JIANG Y F, et al. Independent pilots versus shared pilots: short frame structure optimization for heterogeneous-traffic URLLC networks[J]. IEEE Transactions on Wireless Communications, 2022, 21(8): 5755-5769.

[5] LIN X, LIU A J, HAN C, et al. Joint pilot spacing and power optimization scheme for nonstationary wireless channel: a deep reinforcement learning approach[J]. IEEE Wireless Communications Letters, 2023, 12(3): 540-544.

[6] RAO R M, MAROJEVIC V, REED J H. Adaptive pilot patterns for CA-OFDM systems in nonstationary wireless channels[J]. IEEE Transactions on Vehicular Technology, 2018, 67(2): 1231-1244.

[7] LI F, SHEN B W, GUO J L, et al. Dynamic spectrum access for Internet-of-things based on federated deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2022, 71(7): 7952-7956.

[8] CHEN M, LIU A, LIU W, et al. RDRL: a recurrent deep reinforcement

learning scheme for dynamic spectrum access in reconfigurable wireless networks[J]. IEEE Transactions on Network Science and Engineering, 2021, 9(2): 364-376.

[9] HAN H, XU Y, JIN z, et al. primary-user-friendly dynamic spectrum anti-jamming access: a GAN-enhanced deep reinforcement learning Approach[J]. IEEE Wireless Communications Letters, 2021, 11(2): 258-262.

[10] LI J, GAO H, LV T, et al. Deep reinforcement learning based computation offloading and resource allocation for MEC[C]//Proceedings of IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2018: 1-6.

[11] YANG H, WEI Z, FENG Z, et al. Intelligent computation offloading for MEC-based cooperative vehicle infrastructure system: a deep reinforcement learning approach[J]. IEEE Transactions on Vehicular Technology, 2022, 71(7): 7665-7679.

[12] WANG J, ZHAO L, LIU J, et al. Smart resource allocation for mobile edge computing: a deep reinforcement learning approach[J]. IEEE Transactions on Emerging Topics in Computing, 2019, 9(3): 1529-1541.

[13] NGUYEN T T, REDDI V J. Deep reinforcement learning for cyber security[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(8): 3779-3795.

[14] ZHANG Y, MOU Z, GAO F, et al. UAV-enabled secure communications by multi-agent deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2020, 69(10): 11599-11611.

[15] OH M S, HOSSEINALIPOUR S, KIM T, et al. Channel estimation via successive denoising in MIMO OFDM systems: a reinforcement learning approach[C]//Proceedings of IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2021: 1-6.

[16] CHU M, LIU A, LAU V K N, et al. Deep reinforcement learning based end-to-end multiuser channel prediction and beamforming[J]. IEEE Transactions on Wireless Communications, 2022, 21(12): 10271-10285.

[17] HUANG R, WONG V W S, SCHOBER R. Throughput optimization for grant-free multiple access with multiagent deep reinforcement learning[J]. IEEE Transactions on Wireless Communications, 2020, 20(1): 228-242.

[18] WANG Z, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning[C]//Proceedings of International Conference on Machine Learning. New York: PMLR, 2016: 1995-2003.

[19] European Telecommunications Standards Institute. 3GPP: TS36.104[S]. [2023-05-06].

[作者简介]

刘乔寿（1979− ），男，云南曲靖人，博士，重庆邮电大学副教授、硕士生导师，主要研究方向为 5G 超密集网络干扰协调、云边协同智能计算、FPGA 智能算法加速、物联网系统及终端设备开发。

周雄（1999− ），男，湖北荆州人，重庆邮电大学硕士生，主要研究方向为基于深度强化学习的信道估计。

刘爽（2000− ），男，安徽桐城人，重庆邮电大学硕士生，主要研究方向为空中计算。

邓义锋（1999− ），男，四川成都人，重庆邮电大学硕士生，主要研究方向为空中计算。