

文章编号:1671-251X(2021)06-0109-07

DOI:10.13272/j.issn.1671-251x.2021040053

# 基于 Q-学习算法的矿井自适应 OFDM 调制研究

朱静茹, 张育芝, 王安义, 李萍

(西安科技大学 通信与信息工程学院, 陕西 西安 710054)



扫码移动阅读

**摘要:**针对传统基于固定信噪比门限的自适应 OFDM(正交频分复用)调制技术应用于复杂矿井信道时,由于反馈信道状态与实际信道状态不能完全匹配,导致误码率高和吞吐量低的问题,提出了一种基于 Q-学习算法的自适应 OFDM 调制方法,并将其应用于矿井自适应 OFDM 调制系统。该系统由发送端、矿井无线信道和接收端组成,发送端为矿井下装有传感器的小车,可以在狭长的巷道内自由移动。发送端利用 Q-学习算法在与矿井无线信道的动态交互中不断更新状态-动作值函数,并根据更新的状态-动作值函数,采用贪婪策略来选择调制方式,逼近最优自适应调制策略,以达到降低系统误码率、提高通信吞吐量的目的。与基于 SARSA 算法、固定信噪比门限的 2 种矿井自适应 OFDM 调制系统性能进行仿真对比,结果表明:矿井小车在匀速和移动速度变化状态下,基于 Q-学习算法的自适应 OFDM 调制系统平均误码率分别为  $1.1 \times 10^{-3}$ ,  $2.1 \times 10^{-3}$ ,总吞吐量分别为 3 115 bit, 2 719 bit,均优于基于 SARSA 算法和固定信噪比门限的自适应 OFDM 调制系统,且系统中 Q-学习算法收敛速度优于 SARSA 算法。

**关键词:**矿井无线通信;信道自适应调制;正交频分复用;强化学习;Q-学习算法;SARSA 算法;OFDM

中图分类号:TD655 文献标志码:A

Q-learning algorithm based mine adaptive OFDM modulation

ZHU Jingru, ZHANG Yuzhi, WANG Anyi, LI Ping

(College of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an 710054, China)

**Abstract:** When the traditional adaptive OFDM (Orthogonal Frequency Division Multiplexing) modulation technology based on fixed signal-to-noise ratio threshold is applied to complex mine channels, the feedback channel state cannot completely match the actual channel state, resulting in high bit error rate and low throughput. In order to solve the above problem, a Q-learning algorithm based mine adaptive OFDM modulation method is proposed and applied to the mine adaptive OFDM modulation system. The system is composed of a transmitter, a mine wireless channel and a receiver. The transmitter is a sensor-equipped mine cart, which can move freely in a narrow roadway. The transmitter uses Q-learning algorithm to update the state-action value function continuously in the dynamic interaction with the mine wireless channel. And the transmitter uses a greedy strategy to select the modulation method according to the updated state-action value function to approximate the optimal adaptive modulation strategy so as to reduce the system BER and improve the communication throughput. The performance of two mine adaptive OFDM modulation systems based on SARSA algorithm and fixed signal-to-noise ratio threshold is compared. The result shows that the average BER of the adaptive OFDM modulation system based on

收稿日期:2021-04-16;修回日期:2021-06-20;责任编辑:张强,郑海霞。

基金项目:国家重点研发计划项目(2018YFC0808301);陕西省教育厅科研计划项目(18JK0499)。

作者简介:朱静茹(1995-),女,陕西西安人,硕士研究生,主要研究方向为复杂环境下无线通信,E-mail:19207205075@stu.xust.edu.cn。

引用格式:朱静茹,张育芝,王安义,等.基于 Q-学习算法的矿井自适应 OFDM 调制研究[J].工矿自动化,2021,47(6):109-115.

ZHU Jingru, ZHANG Yuzhi, WANG Anyi, et al. Q-learning algorithm based mine adaptive OFDM modulation [J]. Industry and Mine Automation, 2021, 47(6): 109-115.

Q-learning algorithm are  $1.1 \times 10^{-3}$ ,  $2.1 \times 10^{-3}$ , and the total throughput are 3, 115 bit, 2, 719 bit respectively in the uniform and non-uniform movement states of mine cart. These results are better than the adaptive OFDM modulation system based on SARSA algorithm and fixed signal-to-noise ratio threshold. And the convergence speed of Q-learning algorithm in the system is better than that of SARSA algorithm.

**Key words:** mine wireless communication; channel adaptive modulation; orthogonal frequency division multiplexing; reinforcement learning; Q-learning algorithm; SARSA algorithm; OFDM

## 0 引言

煤矿无线通信系统已成为煤矿信息化和安全生产的重要组成部分。与地面无线通信相比,煤矿井下粗糙的巷壁和障碍物会导致煤矿无线通信系统信号幅度大范围波动,造成信号严重衰落,导致煤矿无线通信系统误码率上升,降低了通信质量<sup>[1]</sup>。因此,如何缓解矿井无线信道衰落对矿井无线通信系统性能的影响是亟需解决的问题。国内外学者针对该问题进行了大量研究,采取自适应 OFDM (Orthogonal Frequency Division Multiplexing, 正交频分复用) 调制技术来对抗衰落,提高频谱利用率,降低误码率。文献[2]将 OFDM 自适应调制技术应用于井下无线多媒体监测场景来缓解频率选择性衰落带来的影响,提高了井下多媒体传感器网络的可靠性。文献[3]将 OFDM 自适应调制技术引入井下无线通信系统,解决了煤矿井下多径干扰严重的问题,提高了井下通信的抗噪性和实时性。文献[4]设计了一种能根据信道特征动态改变其比特和功率分配的自适应 OFDM 算法,提高了井下无线数据传输的抗干扰性。但以上矿井自适应 OFDM 调制技术均假设信道状态已知,依靠提前确定的固定信噪比门限选择相应的调制方式,调制方式不够智能化,且在实际矿井无线信道中由发送端到达接收端的信号是时变信号,信号传输过程中存在时延,且接收端信噪比估计不是理想估计,导致信号发送端接收到的反馈信道状态与实际信道状态不能完全匹配,造成传统基于固定信噪比门限的自适应 OFDM 调制系统误码率高及吞吐量低等问题。

Q-学习算法在一定条件下只需采用贪婪策略即可保证收敛,是一种与模型无关的强化学习算法。针对上述问题,本文提出了一种基于 Q-学习算法的自适应 OFDM 调制方法,并将其用于矿井自适应 OFDM 调制系统。发送端可不用提前已知信道状态信息,在与信道的交互中不断更新状态-动作值函数,最终利用状态-动作值函数学习出动态矿井无线衰落信道下的最佳自适应调制策略,达到降低矿井无线通信系统误码率,提高吞吐量的目的。

## 1 矿井自适应 OFDM 调制系统

### 1.1 系统原理

矿井自适应 OFDM 调制系统由发送端、矿井无线信道和接收端组成,如图 1 所示。发送端为矿井下装有传感器的小车,可以在狭长的巷道内自由移动。

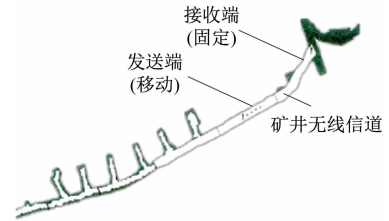


图 1 矿井自适应 OFDM 调制系统组成  
Fig. 1 Composition of mine adaptive OFDM modulation system

矿井小车在移动过程中的自适应调制过程如图 2 所示。发送端在  $B_1$  处向接收端发送信道状态信息(Channel State Information, CSI)测试信号,接收端在经过时延后收到 CSI 测试信号,并对接收到的 CSI 测试信号进行估计后将结果反馈给发送端,最后发送端移动到  $B_3$  处,根据接收到的反馈 CSI 测试信号自适应地分配调制方式。

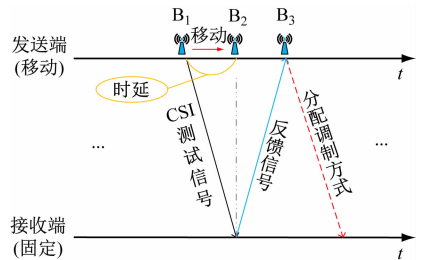


图 2 矿井信道自适应 OFDM 调制过程  
Fig. 2 Adaptive OFDM modulation process of mine channel

### 1.2 矿井无线信道衰落模型

矿井无线信道易受噪声等外界因素影响,存在衰落现象。一般可将巷道内无线信道衰落分为大尺度衰落和小尺度衰落 2 种。

#### 1.2.1 巷道内大尺度衰落

巷道内大尺度衰落包括自由空间路径损耗和矿井环境下特定的电磁波传播损耗<sup>[5-6]</sup>。对于任意距离,特定时间与位置下的路径损耗  $\zeta(t, d)$  服从正态

分布,单位为 dB。

$$\zeta(t, d) = \zeta(t, d_0) + 10n \lg \frac{d}{d_0} + X_\sigma(t) \quad (1)$$

式中: $t$ 为时间; $d$ 为信号传输距离; $d_0$ 为参考距离; $n$ 为路径损耗指数,表示路径损耗增长速率; $X_\sigma(t)$ 为阴影衰落,服从均值为0、方差为 $\sigma^2$ 的正态分布。

矿井巷道中电磁波传播损耗与巷道壁的粗糙度、倾斜度和天线的极化方式等相关,其中垂直极化衰减损耗为

$$L_{\text{ver}} = 4.343\lambda^2 z \left( \frac{\omega^2}{a^3 \sqrt{\varphi_1 - 1}} + \frac{\rho^2 \varphi_2}{b^3 \sqrt{\varphi_2 - 1}} \right) \quad (2)$$

式中: $\lambda$ 为电磁波信号波长; $z$ 为收发天线距离; $\omega$ 为水平传播方向的半波数; $a, b$ 分别为巷道的宽和高; $\varphi_1, \varphi_2$ 分别为两侧和顶底板的相对介电常数; $\rho$ 为垂直传播方向的半波数。

水平极化波的总损耗为

$$L_{\text{total}} = L_{\text{hor}} + L_{\text{rough}} + L_{\text{tile}} \quad (3)$$

式中: $L_{\text{hor}}$ 为水平极化损耗; $L_{\text{rough}}$ 为粗糙损耗; $L_{\text{tile}}$ 为倾斜损耗。

$$L_{\text{hor}} = 4.343\lambda^2 z \left( \frac{\omega^2 \varphi_1}{a^3 \sqrt{\varphi_1 - 1}} + \frac{\rho^2}{b^3 \sqrt{\varphi_2 - 1}} \right) \quad (4)$$

$$L_{\text{rough}} = 8.636\pi^3 \Delta h^2 z \lambda \left( \frac{1}{a^4} + \frac{1}{b^4} \right) \quad (5)$$

$$L_{\text{tile}} = 4.343\pi^2 \vartheta^2 z \frac{1}{\lambda} \quad (6)$$

式中: $\Delta h$ 为巷道表面的起伏高度; $\vartheta$ 为顶壁和巷道壁的倾斜角度。

### 1.2.2 巷道内小尺度衰落

本文以矿井无线信道普遍采用的 Nakagami-m 衰落作为巷道内小尺度衰落模型,其概率密度函数<sup>[7]</sup>为

$$F(r) = \frac{2m^m r^{2m-1}}{\Gamma(m)\Omega^m} \exp(- (m/\Omega)r^2) \quad (7)$$

式中: $r$ 为 Nakagami-m 包络; $m$ 为衰落因子,表征信号衰落强度, $m$ 值越大表示信号衰落越小; $\Gamma(\cdot)$ 为 Gamma 函数; $\Omega$ 为平均功率。

用 Nakagami-m 分布能够准确地表征多径信号在复杂场景的衰落情况。采用正弦求和法仿真 Nakagami-m 衰落信号,将随机序列分解为整数部分和小数部分, Nakagami-m 衰落信号为

$$g(t) = \sqrt{\alpha \sum_{k=1}^p g_1^2(t) + \beta g_2^2(t)} \quad (8)$$

式中: $\alpha$ 和 $\beta$ 分别为调整整数部分和小数部分的参数, $\beta = 2m - \alpha p$ ;  $p$ 为 $2m$ 的整数部分; $g_1(t), g_2(t)$ 为均值为0、方差相等的2个不相关的实高斯随机过程。

采用多个正弦波的叠加来逼近随机过程 $g_1(t)$ 和 $g_2(t)$ ,用等间隔方法来获得多普勒系数和多普勒频率。

$$g_l(t) = \sum_{l=1}^N c_l \cos(2\pi f_l t + \theta_l) \quad (9)$$

式中: $N$ 为正弦波个数; $c_l$ 为第 $l$ 个正弦波的幅度; $f_l$ 为第 $l$ 个正弦波的频率; $\theta_l$ 为第 $l$ 个正弦波的相位。

$$c_l = \left[ \frac{4}{\pi} \left( \arcsin \frac{1}{N} - \arcsin \frac{l-1}{N} \right) \right]^{\frac{1}{2}} \quad (10)$$

$$f_l = \frac{v}{\lambda} \frac{1}{2N} (2l-1) \quad (11)$$

式中: $v$ 为矿井小车运动速度。

### 1.2.3 矿井无线信道衰落特性分析

发送端天线采用水平极化方式, $v = 15$  km/h,  $n = 2, a = 4$  m,  $b = 3$  m,  $\varphi_1 = \varphi_2 = 4.5, \omega = 1, \rho = 1, m = 1, N = 40, d_0 = 1$  m,  $X_\sigma(t)$ 服从均值为0、方差为5的正态分布, $\Delta h$ 服从均值为0.1、方差为0.05的正态分布。基于以上参数对矿井无线信道衰落特性进行分析。矿井无线信道衰落特性如图3所示,蓝色曲线为大尺度衰落和小尺度衰落叠加后的结果,总体呈下降趋势;红色曲线为拟合结果,表征了大尺度衰落的趋势,即收发端距离越远,衰落越严重。

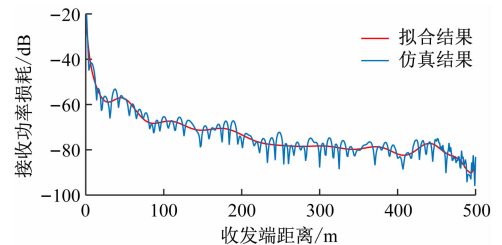


图3 矿井无线信道衰落特性

Fig. 3 Fading characteristics of mine wireless channel

矿井巷道长且不平整,井下设备多,影响井下移动设备的移动速度,不同移动速度对应不同的衰落程度,因此,本文对2种矿井小车移动情况下的衰落特性进行分析。

(1) 假设矿井小车以15 km/h速度从距发送端100~200 m之间往复移动,总路程为1 km,用时240 s,匀速移动状态下矿井信道衰落特性如图4所示。随着矿井小车往复移动,衰落呈现周期性变化。

(2) 矿井小车从距发送端100~200 m之间往复移动,总路程为1 km,前500 m移动速度为15 km/h,后500 m移动速度变为20 km/h,总用时210 s,矿井小车移动速度变化状态下矿井无线信道衰落特性如图5所示。与图4相比,后500 m的信道衰落更加剧烈,这是因为随着移动速度增大, Nakagami-m 衰落更加剧烈。

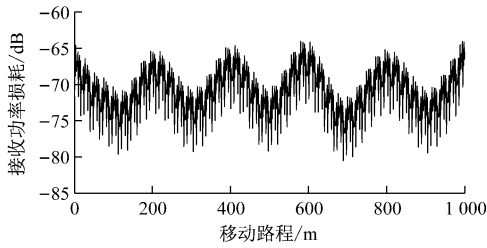


图 4 矿井小车匀速移动状态下矿井无线信道衰落特性  
Fig. 4 Fading characteristics of mine wireless channel as mine cart in the uniform movement state

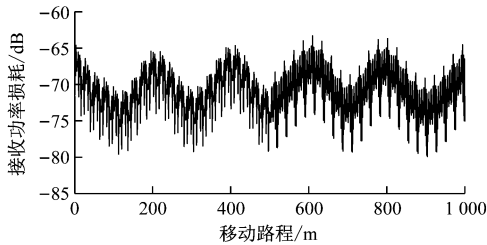


图 5 矿井小车移动速度变化下矿井无线信道衰落特性  
Fig. 5 Fading characteristic of mine wireless channel as mine cart in the non-uniform movement state

## 2 基于 Q-学习算法的自适应 OFDM 调制方法

### 2.1 Q-学习算法基本原理

Q-学习算法是 Chris Watkins 将时间差分与最优控制问题融合后提出的一种强化学习算法,通过对状态-动作值函数进行估计来逼近最优策略。

Q-学习算法框架由智能体、环境、状态、动作和奖励 5 个要素组成,如图 6 所示。在每个时刻,智能体都能从环境中接收到一个状态并通过策略产生一个动作与环境继续交互,环境会产生新的状态,同时给出一个奖励,智能体更新状态-动作值函数并进入下一状态。如此循环,智能体与环境不停交互,从而产生更多的数据(状态和奖励),并利用新的数据进一步改善智能体所采取的动作,最终学习相应任务下的最优策略<sup>[8-9]</sup>。

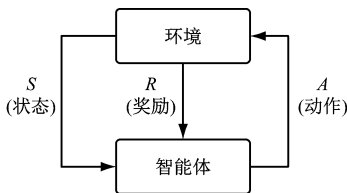


图 6 Q-学习算法原理

Fig. 6 Principle of Q-learning algorithm

Q-学习算法中状态-动作值函数的迭代公式为<sup>[10]</sup>

$$Q(S, A) \leftarrow Q(S, A) + \eta [R + \gamma \max_{A'} Q(S_{next}, A') - Q(S, A)] \quad (12)$$

式中:  $Q(S, A)$  为当前时刻的状态-动作值,  $S$  为当

前时刻的状态,  $A$  为当前时刻所采取的动作;  $\eta$  为学习率;  $R$  为当前时刻所采取动作得到的奖励;  $\gamma$  为折扣因子;  $Q(S_{next}, A_{next})$  为下一时刻的状态-动作值,  $S_{next}$  为下一时刻的状态,  $A_{next}$  为下一时刻所采取的动作。

### 2.2 基于 Q-学习算法的自适应 OFDM 调制

Q-学习算法中智能体为发送端;环境为矿井无线信道;状态集为矿井小车移动过程中不同时刻收到的反馈状态信息即信噪比(SNR);动作集为 OFDM 系统下 BPSK、4QAM、8QAM、16QAM 4 种调制方式;奖励为 4 种调制方式在不同信噪比条件下的吞吐量。

基于 Q-学习算法的矿井自适应 OFDM 调制根据不断更新的状态-动作值函数,采用贪婪策略来分配调制方式,具体步骤<sup>[11-13]</sup>为

步骤 1: 随机初始化 Q 表(一般为 0)。

步骤 2: 初始化信道状态  $S$ , 使用贪婪策略根据信道状态  $S$  选取一个动作  $A$ , 即为选取一个调制方式。

步骤 3: 执行动作  $A$ , 观测奖励值  $R$  和下一个状态  $S_{next}$ 。

步骤 4: 利用式(12)更新 Q 表。

步骤 5: 设置下一状态为当前状态, 即  $S = S_{next}$ , 若  $S$  未达到目标状态, 则转至步骤 3。

根据实际信道状态计算奖励值  $R$ , 贪婪策略以  $\epsilon$  的概率随机选择未知的一个调制方式, 避免系统陷入局部最优,  $1 - \epsilon$  的概率选择当前状态下最大状态-动作值对应的调制方式; 状态-动作值更新函数学习率控制连续时刻所能获得状态-动作值之间的差异, 折扣因子  $\gamma$  决定下一时刻所获得反馈奖励的大小。因此, 合适的贪婪系数  $\epsilon$ 、学习率  $\eta$ 、折扣因子  $\gamma$  是提升整个矿井自适应 OFDM 调制系统性能的关键。

## 3 仿真分析

应用 Matlab 仿真软件, 在矿井小车 2 种移动状态下, 对基于 Q-学习算法的自适应 OFDM 调制系统和基于 SARSA (State - Action - Reward - State - Action) 算法<sup>[14-15]</sup>、固定信噪比门限的自适应 OFDM 调制系统的误码率和吞吐量性能进行仿真对比。

仿真环境如图 1 所示, 接收端在矿井巷道入口的固定位置接收 CSI 信号, 发送端处于移动状态下进行信号发送; 以 4G 通信系统为参考, 将 OFDM 系统载波数设为 1 024 个, 采用 BPSK、4QAM、8QAM、16QAM 4 种调制方式进行自适应 OFDM 调制。

固定信噪比门限值从 OFDM 系统下 4 种调

制方式吞吐量随信噪比变化曲线中得到:当  $SNR < 3$  dB 时选择 BPSK;  $3$  dB  $\leq SNR < 8$  dB 时选择 4QAM;  $8$  dB  $\leq SNR < 22$  dB 时选择 8QAM;  $SNR \geq 22$  dB 时选择 16QAM。Q-学习和 SARSA 算法中  $\epsilon = 0.1, \eta = 0.9, \gamma = 0.9$ 。

2种移动状态对应的衰落信道信噪比序列为发射功率减去对应的接收功率损耗和加性高斯白噪声后得到的序列。其中发射功率为 90 dB,加性高斯白噪声的方差是从 7 dB 的接收噪声数值计算中得到<sup>[16]</sup>。

### 3.1 Q-学习算法参数

在矿井小车匀速移动对应的衰落信道下,仿真  $\epsilon, \eta$  和  $\gamma$  取不同值时对基于 Q-学习算法和 SARSA 算法的自适应 OFDM 调制系统吞吐量性能的影响,  $\epsilon, \eta$  和  $\gamma$  的取值范围为  $(0, 1)$ , 步长为 0.1, 仿真结果如图 7—图 9 所示。

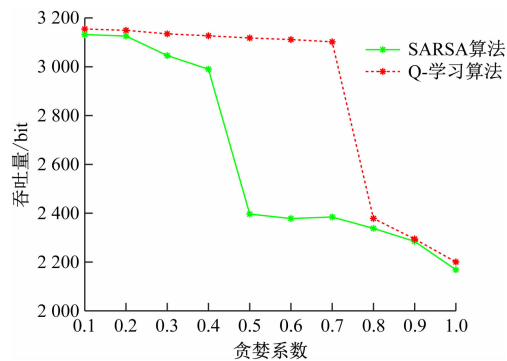


图 7 不同贪婪系数  $\epsilon$  下的系统吞吐量性能

Fig. 7 System throughput performance under different  $\epsilon$

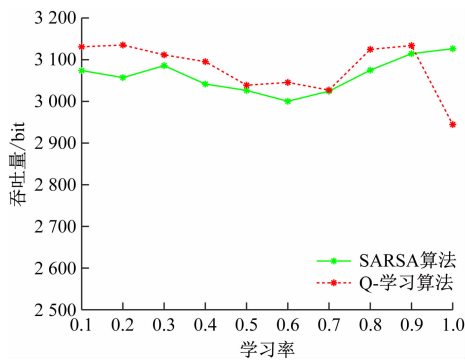


图 8 不同学习率  $\eta$  下的系统吞吐量性能

Fig. 8 System throughput performance under different  $\eta$

从图 7 可看出,随着贪婪系数  $\epsilon$  的增大,基于 Q-学习算法和 SARSA 算法的矿井自适应 OFDM 调制系统吞吐量性能逐渐下降,在贪婪系数取 0.1 时,2 种系统吞吐量性能均达到最佳。从图 8 可看出,在学习率为 0.9 时,基于 Q-学习算法的矿井自适应 OFDM 调制系统取得最大吞吐量。从图 9 可看出,在折扣因子取 0.9 时,基于 Q-学习算法和 SARSA 算法的矿井自适应 OFDM 调制系统均可获得最大吞吐量。

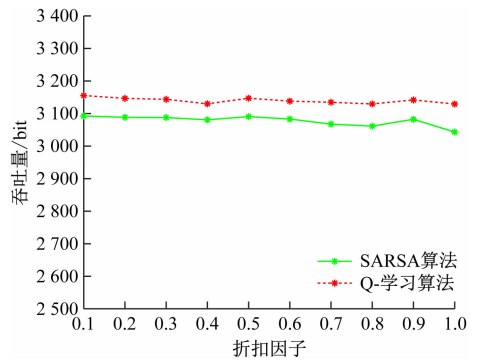


图 9 不同折扣因子  $\gamma$  下的系统吞吐量性能

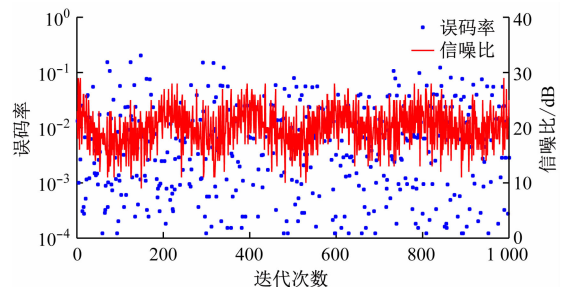
Fig. 9 System throughput performance under different  $\gamma$

综上,当  $\epsilon = 0.1, \eta = 0.9, \gamma = 0.9$  时,基于 Q-学习算法和 SARSA 算法的矿井自适应 OFDM 调制系统均可获得良好的吞吐量性能。

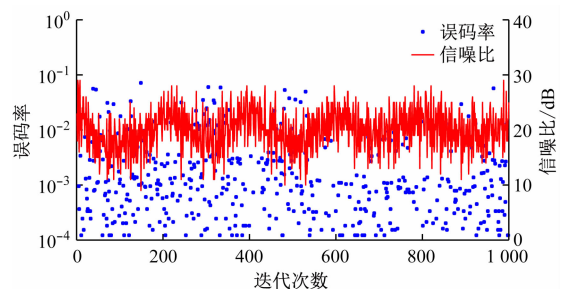
### 3.2 仿真结果

#### 3.2.1 矿井小车匀速移动时仿真结果

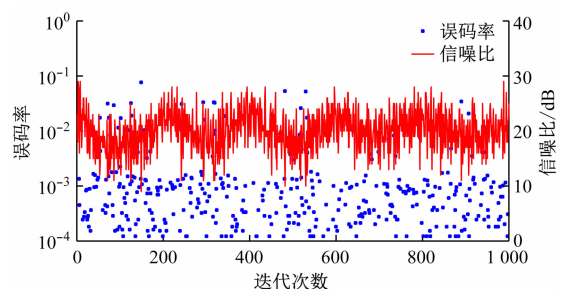
在矿井小车匀速移动情况下,基于固定信噪比门限、SARSA 算法、Q-学习算法的 3 种自适应 OFDM 调制系统性能对比如图 10 所示。



(a) 基于固定信噪比门限的自适应 OFDM 调制系统性能



(b) 基于 SARSA 算法的自适应 OFDM 调制系统性能



(c) 基于 Q-学习算法的自适应 OFDM 调制系统性能

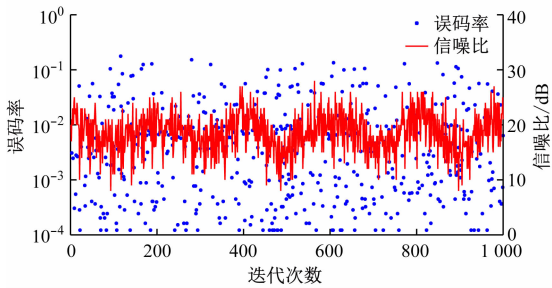
图 10 矿井小车匀速移动时自适应 OFDM 调制系统性能对比

Fig. 10 Comparison of performance of adaptive OFDM modulation system as mine cart in the uniform move state

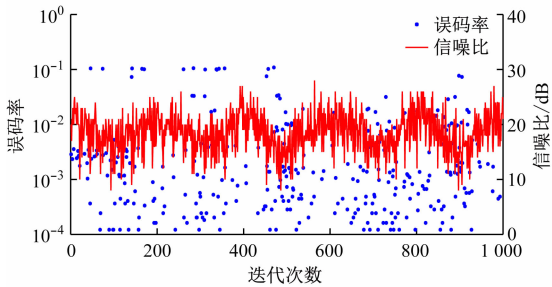
从图 10 可看出,在矿井小车匀速运动情况下,基于 Q-学习算法的自适应 OFDM 调制系统的误码率小于基于固定信噪比门限和 SARSA 算法的自适应 OFDM 调制系统,且误码率在经过多次迭代学习后下降并最终趋于稳定。

### 3.3.2 矿井小车移动速度变化时仿真结果

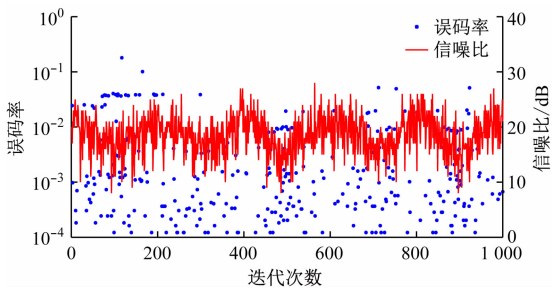
在矿井小车移动速度变化情况下,基于固定信噪比门限、SARSA 算法、Q-学习算法的自适应 OFDM 调制系统性能对比如图 11 所示。



(a) 基于固定信噪比门限的自适应 OFDM 调制系统性能



(b) 基于 SARSA 算法的自适应 OFDM 调制系统性能



(c) 基于 Q-学习算法的自适应 OFDM 调制系统性能

图 11 矿井小车移动速度变化时自适应 OFDM 调制系统性能对比

Fig. 11 Comparison of performance of adaptive OFDM modulation system as mine cart in the non-uniform movement state

从图 11 可看出,在移动速度变化情况下,基于 SARSA 算法和基于 Q-学习算法的自适应 OFDM 调制系统误码率明显小于基于固定信噪比门限的自适应 OFDM 调制系统;基于 SARSA 算法的自适应 OFDM 调制系统在经过 500 次迭代后误码率降低并最终趋于稳定,基于 Q-学习算法的自适应 OFDM 调制系统在经过 300 次迭代后误码率下降并逐渐趋于稳定,其收敛速度优于 SARSA 算法。

对矿井小车 2 种不同移动状态下基于固定信噪比门限、SARSA 算法、Q-学习算法的 3 种自适应 OFDM 调制系统的平均误码率和总吞吐量进行统计,结果见表 1。

表 1 3 种自适应 OFDM 调制系统的平均误码率和总吞吐量

Table 1 Average bit error rate and throughput of three adaptive OFDM modulation systems

调制系统	平均误码率	总吞吐量/bit
基于 Q-学习算法的自适应 OFDM 调制系统(匀速)	$1.1 \times 10^{-3}$	3 115
基于 SARSA 算法的自适应 OFDM 调制系统(匀速)	$2.1 \times 10^{-3}$	3 085
基于固定门限自适应 OFDM 调制系统(匀速)	$6.2 \times 10^{-3}$	2 994
基于 Q-学习算法的自适应 OFDM 调制系统(速度变化)	$2.1 \times 10^{-3}$	2 719
基于 SARSA 算法的自适应 OFDM 调制系统(速度变化)	$2.9 \times 10^{-3}$	2 700
基于固定门限自适应 OFDM 调制系统(速度变化)	$6.9 \times 10^{-3}$	2 670

由表 1 可知,矿井小车匀速移动状态下基于 Q-学习算法的自适应 OFDM 调制系统平均误码率为  $1.1 \times 10^{-3}$ ,总吞吐量为 3 115 bit,性能最佳;矿井小车移动速度变化情况下基于 Q-学习算法的自适应 OFDM 调制系统平均误码率为  $2.1 \times 10^{-3}$ ,总吞吐量为 2 719 bit,性能最佳。

## 4 结论

提出了一种基于 Q-学习算法的矿井自适应 OFDM 调制方法,并将其应用于矿井自适应 OFDM 调制系统。与基于 SARSA 算法和固定信噪比门限的矿井自适应 OFDM 调制系统性能进行了仿真对比,仿真结果表明:

(1) 通过优化 Q-学习算法中贪婪系数、学习率和折扣因子,可以使基于 Q-学习算法的矿井自适应 OFDM 调制系统吞吐量性能达到最佳。

(2) 矿井小车在匀速和移动速度变化状态下,基于 Q-学习算法的自适应 OFDM 调制系统的平均误码率和总吞吐量均优于基于 SARSA 算法和固定信噪比门限的自适应 OFDM 调制系统。

(3) Q-学习和 SARSA 算法应用于矿井自适应 OFDM 调制系统时,Q-学习算法收敛速度优于 SARSA 算法。

## 参考文献(References):

- [1] 魏景新,冉小英. 矿井 OFDM 无线通信系统信道仿真[J]. 辽宁工程技术大学学报(自然科学版),2015,34(12):1345-1349.  
WEI Jingxin, RAN Xiaoying. Channel simulation underground OFDM wireless communication system[J]. Journal of Liaoning Technical University(Natural Science),2015,34(12):1345-1349.
- [2] 范新越,杨维. 井下多媒体无线传感器网络自适应调制研究[J]. 煤炭学报,2009,34(9):1291-1296.  
FAN Xinyue, YANG Wei. Study on the adaptation modulation of underground mine multimedia wireless sensor network[J]. Journal of China Coal Society, 2009,34(9):1291-1296.
- [3] 王娟,付家才. 煤矿井下 OFDM 自适应调制技术[J]. 黑龙江科技学院学报,2010,20(2):135-137.  
WANG Juan, FU Jiakai. OFDM adaptive techniques in coal mine[J]. Journal of Heilongjiang Institute of Science & Technology,2010,20(2):135-137.
- [4] 李宁. 自适应 OFDM 算法在矿井无线通信中的应用研究[J]. 煤矿机电,2015,36(4):12-16.  
LI Ning. Application research on adaptive OFDM algorithm in mine wireless communication [J]. Colliery Mechanical & Electrical Technology, 2015, 36(4):12-16.
- [5] 郭世坤. 基于支持向量机的矿井无线信道建模与精确预测[D]. 西安:西安科技大学,2015.  
GUO Shikun. Wireless channel modeling and accurate prediction in mine based on support vector machine[D]. Xi'an: Xi'an University of Science and Technology,2015.
- [6] 马晶晶. 煤矿复杂环境下的无线信道特性分析与建模仿真[D]. 西安:西安科技大学,2019.  
MA Jingjing. Research and simulation on characteristics of wireless channel in the mine tunnel [D]. Xi'an: Xi'an University of Science and Technology,2019.
- [7] 帅路军,王旭东,宫宇,等. 不同 Nakagami-m 信道仿真方法的性能[J]. 大连海事大学学报,2006,32(3):89-92.  
SHUAI Lujun, WANG Xudong, GONG Yu, et al. Investigation of different Nakagami-m channel simulation methods[J]. Journal of Dalian Maritime University,2006,32(3):89-92.
- [8] 杜威,丁世飞. 多智能体强化学习综述[J]. 计算机科  
学,2019,46(8):1-8.
- DU Wei, DING Shifei. Overview on multi-agent reinforcement learning[J]. Computer Science,2019,46(8):1-8.
- [9] 陈学松,杨宜民. 强化学习研究综述[J]. 计算机应用研究,2010,27(8):2834-2838.  
CHEN Xuesong, YANG Yimin. Reinforcement learning: survey of recent work [J]. Application Research of Computers,2010,27(8):2834-2838.
- [10] JANG B, KIM M, HARERIMANA G, et al. Q-learning algorithms: a comprehensive classification and applications [J]. IEEE Access, 2019, 7: 133653-133667.
- [11] 李程坤. 基于强化学习的自适应调制编码技术的研究[D]. 杭州:杭州电子科技大学,2018.  
LI Chengkun. The adaptive modulation coding technology research based on reinforcement learning [D]. Hangzhou: Hangzhou Dianzi University, 2018.
- [12] QIU Shunan, LI Zhenhai, LI Zhengwei, et al. Model-free optimal chiller loading method based on Q-learning[J]. Science and Technology for the Built Environment,2020,26(8):1100-1116.
- [13] 李萍. 基于强化学习的水声通信自适应调制算法研究[D]. 西安:西安科技大学,2020.  
LI Ping. Research on adaptive modulation algorithm of underwater acoustic communication based on reinforcement learning[D]. Xi'an: Xi'an University of Science and Technology,2020.
- [14] 王安义,李萍,张育芝. 基于 SARSA 算法的水声通信自适应调制[J]. 科学技术与工程,2020,20(16):6505-6509.  
WANG Anyi, LI Ping, ZHANG Yuzhi. Adaptive modulation of underwater acoustic communication based on SARSA algorithm[J]. Science Technology and Engineering,2020,20(16): 6505-6509.
- [15] REN He, DAI Jing, ZHANG Huaguang, et al. Off-policy integral reinforcement learning algorithm in dealing with nonzero sum game for nonlinear distributed parameter systems [J]. Transactions of the Institute of Measurement and Control, 2020, 42(8):2919-2928.
- [16] 陈若山,王艳芬. UWB 定位算法在煤矿井下的应用研究[J]. 工矿自动化,2008,34(6):5-8.  
CHEN Ruoshan, WANG Yanfen. Application research of UWB positioning algorithm of coal mine underground [J]. Industry and Mine Automation, 2008,34(6): 5-8.