

引用本文: 晏丹, 冯楠, 左晓博, 等. 基于深度强化学习的 C+L 波段弹性光网络频谱分配算法[J]. 光通信技术, 2024, 48(3): 23–29.

基于深度强化学习的 C+L 波段弹性光网络频谱分配算法

晏丹^{1,2}, 冯楠^{3,4}, 左晓博^{1,2}, 沈凌飞^{1,2}, 任丹萍^{1,2}, 胡劲华^{1,2}, 赵继军^{1,2*}

(1.河北工程大学 信息与电气工程学院, 河北 邯郸 056038; 2.河北工程大学 河北省安防信息感知与处理重点实验室, 河北 邯郸 056038; 3.中国电子科技集团公司 第五十四研究所, 石家庄 050081; 4.河北省光子信息技术与应用重点实验室, 石家庄 050081)

摘要: 针对 C+L 波段弹性光网络中受激喇曼散射 (SRS) 效应导致物理层损伤加剧的问题, 提出一种基于深度强化学习 (DRL) 自适应调制格式的频谱分配算法, 在路由阶段, 采用 K 最短路由算法为业务请求预计算 K 条最短备选路径; 在波段、调制格式与频谱分配阶段, 采用 DRL 进行智能化决策, 并结合了 2 种奖励函数, 以降低网络阻塞率并提高频谱使用效率。仿真结果表明, 该算法能够有效降低阻塞率并提高频谱利用率。

关键词: C+L 波段弹性光网络; 路由与频谱分配; 受激喇曼散射效应; 深度强化学习; 奖励设计

中图分类号: TN929.1 **文献标志码:** A **文章编号:** 1002-5561(2024)03-0023-07

DOI: 10.13921/j.cnki.issn1002-5561.2024.03.005

开放科学(资源服务)标识码(OSID):



Spectrum allocation algorithm for C+L band elastic optical networks based on deep reinforcement learning

YAN Dan^{1,2}, FENG Nan^{3,4}, ZUO Xiaobo^{1,2}, SHEN Lingfei^{1,2}, REN Danping^{1,2}, HU Jinhua^{1,2}, ZHAO Jijun^{1,2*}

(1. School of Information and Electrical Engineering, Hebei University of Engineering, Handan Hebei 056038, China;

2. Hebei Key Laboratory of Security and Protection Information Sensing and Processing, Hebei University of Engineering,

Handan Hebei 056038, China; 3. The 54th Research Institute of CETC, Shijiazhuang 050081, China;

4. Hebei Key Laboratory of Photonic Information Technology and Application, Shijiazhuang 050081, China)

Abstract: Aiming at the problem of intensified physical layer damage caused by stimulated Raman scattering (SRS) effect in C+L band elastic optical networks, a spectrum allocation algorithm based on deep reinforcement learning (DRL) adaptive modulation format is proposed. In the routing stage, the K -shortest routing algorithm is used to pre calculate K shortest candidate paths for business requests. In the stages of band, modulation format, and spectrum allocation, DRL is used for intelligent decision-making, and two reward functions are combined to reduce network blocking rate and improve spectrum utilization efficiency. The simulation results show that the algorithm can effectively reduce blocking rate and improve spectrum utilization.

Key words: C+L band elastic optical network, routing and spectrum allocation, stimulated Raman scattering effect, deep reinforcement learning, reward design

0 引言

随着增强现实、虚拟现实、元宇宙和 ChatGPT 等

新兴应用的迅猛发展,网络流量呈现出急剧增长的趋势^[1]。为满足不断增长的数据传输需求,多波段弹性光网络(MB-EONs)凭借其无需重新铺设光纤的特性,成为了一种极具经济性和实用性的解决方案。MB-EONs 是一种利用常规 C 波段以外的低损耗波段(O、E、S 和 L 波段,其衰减系数小于 0.4 dB/km 的)进行传输的方案^[2]。其中,由于 L 波段与 C 波段的损耗相近,并且相应的放大器、光交叉连接器等器件相对成熟,因此以 L 波段为扩展目标的 C+L 波段弹性光网络(C+L-EONs)获得了工业界和学术界的高度关注。然而,引入 L 波段也加剧了受激喇曼散射(SRS)效应,在该效应与其它物理层损伤因素的共同作用下,信号在传输中受到

收稿日期: 2024-01-02。

基金项目: 河北省硕士在读研究生创新能力培养资助项目(CXZ-ZSS2024101)资助。

作者简介: 晏丹(2000—),女,硕士研究生,现就读于河北工程大学信息与电气工程学院信息与通信工程专业,研究方向为光纤通信与宽带通信网,主要从事基于机器学习的弹性光网络路由选择与频谱分配相关研究工作,参与了基于深度强化学习的 C+L 波段弹性光网络频谱分配算法研究项目。

* 通信作者: 赵继军(1970—),男,博士,教授,主要研究方向为光纤通信与宽带通信网。



晏丹, 冯楠, 左晓博, 等: 基于深度强化学习的 C+L 波段弹性光网络频谱分配算法

的非线性干扰(NLI)增加,造成了业务传输质量(QoT)的劣化,严重时甚至导致信号在接收端无法被正确解码和处理^[3]。此外,相关信道上 NLI 的严重程度与频谱资源分配决策密切相关,而合理的频谱分配策略可以有效减轻 NLI 对业务 QoT 的负面影响^[4]。为了有效保证业务请求的 QoT,同时实现资源的高效利用,需要设计一个合理可靠的路由、波段、调制格式和频谱分配(RBMSA)算法。现有的 RBMSA 算法主要采用基于固定规则的启发式算法。文献[5]在 C+L+S 波段场景下,将业务根据不同的传输距离需求分为 3 类,根据需求特性,提出了一种基于波段分配优先级的 RBMSA 算法,有效提高了频谱利用率。然而,在业务持续时间内,SRS 效应的影响会动态变化。为了全面考虑新业务对已有业务的影响,文献[6]提出了一种基于信噪比重检测的 RBMSA 算法。在为新业务分配频谱资源时,该算法重新计算路径上其它业务的信噪比,以保证原有业务的 QoT。然而,在 C+L-EONs 中,由于业务类型与调制格式多样化,以及物理层损伤与 RBMSA 过程的密切耦合,利用启发式算法进行频谱资源管理变得愈加复杂。

近年来,随着人工智能的快速发展,人们通过机器学习算法可以更高效地控制 RBMSA 问题的决策过程,从而提高频谱资源利用率,并进一步提升网络性能^[7]。文献[8]首次提出了适用于 EON 中路由与频谱分配的深度强化学习(DRL)框架的资源分配算法。该框架以业务相关信息和所有光纤链路上的频谱占用情况作为输入,通过神经网络计算得出每个资源分配方案对应的 Q 值,并最终选择 Q 值最大的方案作为最优策略。与常见的启发式算法相比,该算法虽然能够进一步降低网络的阻塞率,但未考虑 SRS 效应对业务 QoT 的影响,因此不适用于 MB-EONs 中的资源分配。在此基础上,文献[9]聚焦 MB-EONs,在损伤评估过程中特别关注了 SRS 的影响。然而,该研究采用离线评估的方法,这种方法基于网络中最差的光信噪比(OSNR)状况进行设计,忽视了实际网络中的动态变化,从而不利于频谱资源利用率的提升。相比之下,文献[10]提出了一种创新的在线 OSNR 评估方法,该方法利用高斯噪声进行仿真,无需对路由预计算,各链路被分配特定数值,智能体在选路后通过计算这些数值总和来评估路径的优劣。该算法采用 DRL 方法优化路由选择,但在实际应用中,由于某些原因,该算法尚未能够根据业务需求自适应选择频谱分配和调制格式。

综上所述,为了在 C+L-EONs 中保证业务 QoT 的同时提升频谱资源利用率,本文以置信域策略梯度优化(TRPO)为基础,提出一种基于 DRL 的自适应调制格式(AMF)C+L+EONs 频谱资源分配算法(下文简称 AMF-TRPO 算法)。

1 问题描述

1.1 网络模型

本文将 C+L-EONs 中的物理拓扑建模为一个有向辅助图 $G(V, E, F)$, 其中 V 和 E 分别表示物理拓扑中节点和光纤链路的集合, F 为每条光纤链路上的频隙数量集合。一个动态光路请求可以用 $R(s, d, b, \tau)$ 表示, 其中, s 为源节点, d 为目的节点, b 为带宽, τ 为服务时间。为了保证业务的 QoT, 需要为其计算一条端到端的路由路径, 并选择合适的通信波段和调制格式, 在满足频谱分配三大约束条件(频谱连续性、频谱邻接性和频谱不重叠性)下, 最终完成频谱资源的分配。在整个过程中, 需要优化频谱资源的分配, 以减少物理层损伤的影响, 确保业务的 QoT。

1.2 物理层损伤影响

在 C+L-EONs 中, 除了克尔效应引起的放大器自发辐射(ASE)噪声和 NLI 外, 在 OSNR 分析中还必须考虑通道间的 SRS 效应。SRS 是一种将光信号功率从较低波长通道转移到较高波长通道的非线性效应, 该效应的具体影响如图 1 所示。其中, SRS 效应对业务的影响程度与业务所需带宽及业务间的中心频率间隔有关。C 波段中的业务频隙占用越多, 就有越多的箭头指向 L 波段内的业务, 即 L 波段内的业务受损伤影响越严重。通过对比图 1 中的链路 1 和链路 2 可知, C 波段内链路 2 比链路 1 占用的频隙数多, 因此其 L 波段内业务受到的 SRS 影响越严重; 对比链路 2 和链路 3 可知, 两者 C 波段内占用的频隙数相同, 但链路 3 上

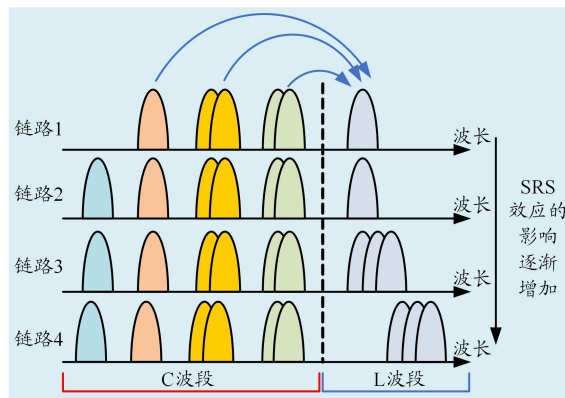


图 1 业务受到的 SRS 效应影响示意图

的 L 波段内业务占用的频隙数更多, 因此链路 3 上的 L 波段内的业务比链路 2 上的业务受到的损伤影响更大。

1.3 物理层损伤评估

为了在传输过程中快速准确地评估业务所受的物理层损伤状况, 本文采用 SRS 的封闭式高斯噪声模型^[11], 旨在对物理层损伤进行准确的量化。假设请求 R 的通信信道为 r , 则其 OSNR 可以由式(1)进行计算。信道 r 对应的信道 r 的累计 ASE 噪声功率 P_{ASE} 和 n 个跨度后的非线性干扰系数 $\eta_n(f_r)$ 计算公式分别如式(2)和式(3)所示。

$$r_{\text{OSNR}} = \frac{P_r}{P_{\text{ASE}} + \eta_n(f_r) P_r^3} \quad (1)$$

$$P_{\text{ASE}} = 2n_{\text{sp}} h f_r B_r (\alpha D - 1) \quad (2)$$

$$\eta_n(f_r) = \sum_{i=1}^n [\eta_{\text{SPM},i}(f_r) + \eta_{\text{XPM},i}(f_r)] \quad (3)$$

其中, P_r 为信道 r 的发射功率, n_{sp} 为自发辐射因子, h 为普朗克常数, f_r 和 B_r 分别为信道 r 的中心频率和带宽, α 为光纤衰减系数, D 为单个跨度的距离。通过式(3)可以得出, $\eta_n(f_r)$ 由第 i 个跨度的自相位调制(SPM)干扰系数 $\eta_{\text{SPM},i}(f_r)$ 和第 i 个跨度的交叉相位(XPM)干扰系数 $\eta_{\text{XPM},i}(f_r)$ 两部分组成。 $\eta_{\text{SPM},i}(f_r)$ 和 $\eta_{\text{XPM},i}(f_r)$ 的计算公式分别如式(4)和式(5)所示。

$$\eta_{\text{SPM},i}(f_r) = \frac{4}{27} \frac{\gamma_1}{B_r^2} \frac{\pi}{\phi_r \alpha^2} \left[\frac{(2\alpha - P_{\text{tot}} C_r f_r)^2 - \alpha^2}{\alpha} \times \right. \\ \left. \text{arcsinh} \left(\frac{\phi_r - B_r^2}{\alpha \pi} \right)^2 + \frac{4\alpha^2 - (2\alpha - P_{\text{tot}} C_r f_r)^2}{2\alpha} \times \right. \\ \left. \text{arcsinh} \left(\frac{\phi_r - B_r^2}{2\alpha \pi} \right) \right] \quad (4)$$

$$\eta_{\text{XPM},i}(f_r) = \frac{32}{81} \sum_{r'=1, r' \neq r}^{N_{s,i}} \left(\frac{P_{r'}}{P_r} \right)^2 \frac{\gamma^2}{B_{r'} \phi_{r,r'} \alpha^2} \times \\ \left[\frac{(2\alpha - P_{\text{tot}} C_r f_{r'})^2 - \alpha^2}{\alpha} \text{arcsinh} \left(\frac{\phi_{r,r'} - B_r^2}{\alpha} \right)^2 + \right. \\ \left. \frac{4\alpha^2 - (2\alpha - P_{\text{tot}} C_r f_r)^2}{2\alpha} \text{arcsinh} \left(\frac{\phi_r - B_r^2}{2\alpha} \right) \right] \quad (5)$$

其中, γ_1 为光纤的非线性系数, B_r 为其它已有业务的带宽。 P_{tot} 和 P_r 分别为所有信道的总发射功率和信道的发射功率, C_r 为归一化喇曼增益谱的线性回归斜率, f_r 为同一条链路上其它已有业务的信道中心频率,

$N_{s,i}$ 为链路 i 上其它业务的数量, r' 表示与当前业务 r 同链路的其它的业务。 ϕ_r 和 $\phi_{r,r'}$ 可通过式(6)和式(7)进行计算。

$$\phi_r = \frac{2}{3} \pi^2 (\beta_2 + 2\pi \beta_3 f_r) \quad (6)$$

$$\phi_{r,r'} = 2\pi^2 (f_r - f_{r'}) [\beta_2 + \pi \beta_3 (f_r + f_{r'})] \quad (7)$$

其中, β_2 和 β_3 分别为二阶群速度色散和三阶群速度色散参数。

综上所述, 业务所受的物理层损伤不仅与当前业务的路由长度、带宽需求 and 对应信道的中心频谱有关, 还与同一条路径上其它已有业务的数量及其相关信道的中心频率紧密相关。基于上述损伤计算模型的评估结果, 本文能够有效判断智能体在调制格式决策方面的优劣。在前向纠错码率为 10^{-3} 时, 每一个可用的调制格式所需的最小 OSNR 阈值^[12]如表 1 所示。此外, NLI 中的 XPM 干扰和当前的请求及未来的请求相关, 但是在分配频谱资源时未来的 XPM 干扰是未知的, 为此, 本文采纳了与文献[13]相符的假设, 全面估算业务在传输过程中受到的 XPM 干扰: 1) 未来请求的数量通过网络中连接请求的平均到达率和当前业务的服务持续时间的乘积进行估计; 2) 未来请求的带宽等于网络中请求的平均带宽; 3) 未来请求使用的频隙数量采用最低调制格式来计算; 4) 未来请求分配的频隙位于当前请求所占用的最高索引频隙的右侧。

表 1 不同调制格式下的频谱效率与 OSNR 阈值

调制格式	频谱效率	OSNR 阈值/dB
BPSK	2	9
QPSK	4	12
8QAM	6	16
16QAM	8	18.6
32QAM	10	21.6

2 AMF-TRPO 算法

2.1 RBMSA 的 DRL 建模

DRL 是机器学习的一种形式, 不同于监督学习和无监督学习, DRL 通过智能体与环境的不断交互 (采取动作) 来获得奖励, 并通过不断优化其动作策略来最大化长期收益 (奖励之和)。由于在动态资源分配过程中无法获取状态转移信息, 因此本文采用数据驱动无模型的方法。具体方法如下: 在业务到达时, 需要将当前业务信息和网络环境信息抽象为状态 S_t , 并输入神经网络, 其中 t 表示时刻 t 。本文使用指数线性单元

晏丹, 冯楠, 左晓博, 等: 基于深度强化学习的 C+L 波段弹性光网络频谱分配算法

作为神经网络的激活函数。神经网络读取状态数据并输出不同 RBMSA 策略的概率分布, 即策略, 智能体根据此策略采取动作 A_t 并尝试建立光路, 奖励系统根据光路建立的结果反馈给智能体相应的奖惩值 R_t 。若有足够的空闲频隙资源分配给业务, 且频谱分配后光路的信噪比可以满足调制格式需求, 则业务能够被成功服务, 此时智能体反馈正值奖励。相反, 智能体将反馈负值奖励。最后, 神经网络会根据反馈的奖励不断更新策略, 以便在后续的动作选择中输出最优动作, 从而获得最大化的累计奖励值。

2.2 TRPO 算法框架

本文采用 TRPO 算法进行训练。在文献[7]中, 该算法在应对不同的网络拓扑时, 相较于其它常见的 DRL 算法(如 DQN、A2C 等), 展现出了更好的鲁棒性。因为 TRPO 算法的最终目标是通过最大化策略的性能, 即最大化累计奖励的期望, 以实现模型的稳健性。策略 π_θ 在环境中执行时, 轨迹的期望累积奖励 $J(\theta)$ 的具体表达式如式(8)所示。

$$J(\theta) = E_{s_0} [V^{\pi_\theta}(s_0)] = E_{\pi_\theta} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \right] \quad (8)$$

其中, E_{s_0} 表示在初始状态分布上的期望值, E_{π_θ} 表示在策略 π_θ 下的期望值, $V^{\pi_\theta}(s_0)$ 是状态价值函数, γ 是折扣因子, 范围在 $[0, 1]$, $r(s_t, a_t)$ 为状态 s_t 下采取动作 a_t 的奖励值。

在使用 TRPO 算法进行学习时, 首先需要使用当前策略与环境进行交互, 收集一批经验样本。这些样本通常包括状态、动作和奖励等信息。接下来, 使用收集到的样本数据计算每个状态下动作的优势函数估计。优势函数表示某个动作相对于平均水平的优势程度, 其表达式如式(9)所示。

$$A_\pi(s, a) = Q_\pi(s, a) - V_\pi(s) = E_{s' \sim P(s'|s, a)} [r(s) + \gamma V^\pi(s') - V^\pi(s)] \quad (9)$$

其中, $Q_\pi(s, a)$ 为状态动作价值函数, $V_\pi(s)$ 与 $V^\pi(s')$ 分别表示位于状态 s 及下一个状态 s' 时的状态价值函数。

基于优势函数计算出的结果, 结合损失函数来更新策略参数, TRPO 算法的损失函数如式(10)所示。

$$L(\theta) = E_{s, a \sim \pi_{\text{old}}} \left[\frac{\pi_\theta(a|s)}{\pi_{\text{old}}(a|s)} A_{\text{old}}(s, a) \right] \quad (10)$$

其中, $\pi_\theta(a|s)$ 为当前策略(由参数 θ 控制)下在状态 s 时采取动作 a 的概率, $\pi_{\text{old}}(a|s)$ 为更新之前的策略在状态 s 时采取动作 a 的概率, $A_{\text{old}}(s, a)$ 为更新之前的策略下选择动作 a 的优势函数。

在策略更新过程中, TRPO 通过限制新旧策略之间的库尔贝克-莱布勒(KL)散度, 控制更新参数的步长, 从而保证算法的稳定性和收敛性。

2.3 状态空间

当请求 $R(s, d, b, t)$ 到达时, 本文算法利用智能体来分配最佳的调制格式。与基于路由长度的距离自适应^[14]不同, 本文通过计算业务传输信道的实际 OSNR 来选择调制格式, 并合理考虑未来业务对当前业务的影响。因此, 状态空间定义为 $1 \times [2 \cdot V + 1 + (2 \cdot J + 3) \cdot M \cdot B \cdot K]$ 大小的矩阵, 该矩阵定义如下:

$$S_i = \left\{ s, d, \tau, \left\{ \left\{ \left\{ \begin{matrix} 1, j \\ z_{k, w, m}^{1, j}, z_{k, w, m}^{2, j} \end{matrix} \right\} \right\}_{j \in [1, J]}, z_{k, w, m}^{3, j}, \right. \right. \\ \left. \left. z_{k, w, m}^{4, j}, z_{k, w, m}^{5, j} \right\}_{m \in [1, M]} \right\}_{w \in [1, 2]} \right\}_{k \in [1, K]} \quad (11)$$

其中, $2 \cdot V$ 代表连接请求的源节点 s 与目的节点 d , 采用 one-hot 编码, 将源节点 s 与目的节点 d 用 1 表示, 其余节点用 0 表示。 J 为可用的频隙块数量, M 为考虑的调制格式总数, B 为所考虑的波段数量, K 为备选的最短路径数量。对于 K 条备选路径中的第 k 条路径, 在波段 w 下使用 M 种不同的调制格式去计算一个请求在调制格式 m 时所需要的频隙数量 $z_{k, w, m}^{3, j} \circ z_{k, w, m}^{1, j}$ 、 $z_{k, w, m}^{2, j}$ 、 $z_{k, w, m}^{4, j}$ 、 $z_{k, w, m}^{5, j}$ 分别表示在第 k 条路径上使用波段 w 和调制格式 m 时可用频隙块 j 中包含的频隙数量、起始频隙索引、所需频隙数量、 J 个频隙块的平均频隙数量和总数量。

2.4 动作空间

当业务请求到达时, 本文算法利用 K 条最短路径(KSP)算法为业务找到 K 条可用的候选链路, 考虑使用 C 或者 L 波段, 先选择一种合适的调制格式来确定其所需频隙数, 再对每条链路找到 J 个可用的候选频隙块。对于每一个连接请求, 智能体都将从动作空间中为其选择一条路径 k , 一个分配波段 w , 一种调制格式 m 和一个频隙块 j , 因此, 动作空间大小为 $K \times B \times M \times J$ 。

2.5 奖励设计

为了反馈智能体的行为, 本文设置了以下奖励机制, 如式(12)所示。当成功服务请求时, R_0 设置为 1, 否则奖励为 -1。其次, 借鉴文献[15-16]将启发式信息与奖励设置相结合的思想, 以提高 DRL 的效率和质量, 如式(15)所示。根据波段的可用性, 设置奖励函数 R_1 , 其值为链路上给定波段上已占用频隙与总频隙数量之比。如果请求被接受, 智能体将获得平均波段空闲率 $\tilde{A}(B)$ 作为奖励; 如果请求被拒绝, 则奖励值为 -1。因

此,奖励的范围是 $-1 \cup [0, 1]$ 。具体设置的细节如下:

$$R_0 = \begin{cases} 1, & \text{业务服务成功} \\ -1, & \text{业务阻塞} \end{cases} \quad (12)$$

$$A_n(B) = \frac{F_{\text{free}}}{F_{\text{total}}} \quad (13)$$

$$\tilde{A}(B) = \frac{1}{N} \sum_{l \in p} A_l(B) \quad (14)$$

$$R_1 = \begin{cases} \tilde{A}(B), & \text{业务服务成功} \\ -1, & \text{业务阻塞} \end{cases} \quad (15)$$

其中, F_{free} 为链路上空闲的频隙数, F_{total} 为链路上总的频隙数, N 为所选路径包含的链路数, $A_n(B)$ 和 $\tilde{A}(B)$ 分别为链路 l 上的波段空闲率及所选路径 p 上波段的平均空闲率。

3 仿真性能与分析

3.1 仿真环境

本文在具有 14 个节点、21 条链路的 NSFNET 网络拓扑^[17]中进行了仿真实验。在仿真过程中,假设每根光纤中 C 波段和 L 波段分别包含 358 和 558 个可用频隙^[18],每个频隙带宽为 12.5 GHz。业务的源、宿节点和请求速率均随机生成,请求速率为 10~400 Gb/s,保护带宽为 2 个频隙。业务到达率服从参数为 λ 的泊松分布,持续时间服从参数为 μ 的负指数分布,即全网的总负载为 λ/μ ,仿真中使用的物理层参数如表 2 所示。每个回合包含 50 个训练步骤,总共进行了 50 万次训练,训练总步数为 5×10^5 步,其中折扣因子设置为 0.95,学习率设置为 10^{-5} 。

3.2 仿真结果分析

当业务量为 2 000 Erlang、候选路径数 K 为 5、候

表 2 物理层仿真参数

仿真参数	值
C 波段和 L 波段自发辐射因子 n_{sp}	1.25 和 1.99
光信号的发射功率 P/dBm	0
普朗克常数 $h/(\text{J} \cdot \text{s})$	$6.626\ 107\ 015 \times 10^{-34}$
光纤衰减系数 $\alpha/(\text{dB/km})$	0.2
归一化喇曼增益谱的线性回归斜率 $C_1/[\text{W}^{-1}/(\text{km} \cdot \text{THz}^{-1})]$	0.028
光纤的非线性系数 $\gamma/(\text{W} \cdot \text{km})^{-1}$	1.2
二阶群速度色散 $\beta_2/(\text{ps}^2/\text{km})$	-21.6
三阶群速度色散参数 $\beta_3/(\text{ps}^3/\text{km})$	0.14

选频隙块 J 为 1 时,所提算法的收敛情况如图 2 所示。其中, $\text{AMF-TRPO-}R_0$ 和 $\text{AMF-TRPO-}R_1$ 分别表示所提算法与 2 种奖励函数的结合。图 2 中,在初始阶段,回合内的累计奖励回合内的累计奖励值较低,反映出由于缺乏先验经验,智能体在资源分配上显得较为随机,特别是在大量业务请求遭遇阻塞的情况下。然而,随着训练的深入和策略网络参数的持续更新,奖励逐步稳定,呈现出更加稳健的状态。此外,2 种奖励函数对应的学习曲线呈现出不同的趋势。 $\text{AMF-TRPO-}R_0$ 的曲线呈现出明显的上升趋势。这是因为在业务被成功服务的情况下给予较大的正向奖励,这种设置使得算法在训练过程中更加注重提高业务服务的成功率,因此曲线的上升趋势明显,斜率较大; $\text{AMF-TRPO-}R_1$ 的曲线则呈现出相对平缓的上升趋势。这是因为在计算波段平均空闲率时,没有明确的正向或负向边界,导致奖励值的变动范围较小,从而使得曲线的上升趋势相对平缓。

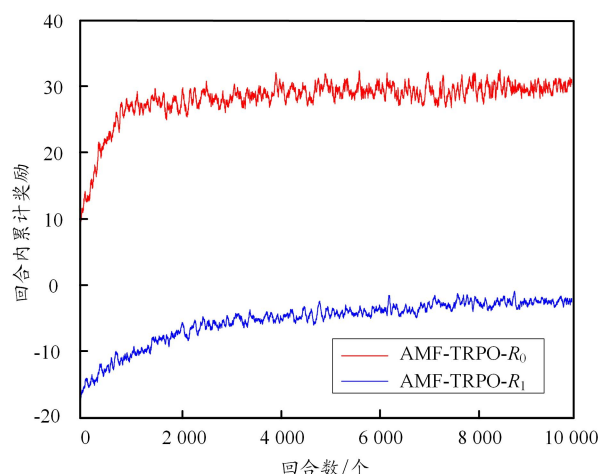


图 2 所提算法收敛效果图

为进一步验证所提算法的性能,在算法模型达到稳定状态后,本文对比了所提算法和基准算法($\text{TRPO-}R_0$ ^[9]与 $\text{TRPO-}R_1$ ^[16])的阻塞率性能,如图 3 所示。其中, $\text{TRPO-}R_0$ 根据距离自适应技术为业务选择合适的调制格式,并利用 DRL 实现 RBMSA 中的路由与频谱分配决策。 $\text{TRPO-}R_1$ 在 $\text{TRPO-}R_0$ 的基础上以波段的可用性作为奖励函数,引导智能体优先选择占用较少频谱的波段。可以看出,当业务量在 1 000~4 000 Erlang 时,随着负载的增加,4 种算法的阻塞率都随之增加。其中, $\text{TRPO-}R_1$ 比 $\text{TRPO-}R_0$ 的阻塞率更低,这是因为 $\text{TRPO-}R_1$ 优先将请求分配到占用较少的频段,提高了为业务成功分配资源的概率。在业务量为 1 000 Erlang

晏丹, 冯楠, 左晓博, 等: 基于深度强化学习的 C+L 波段弹性光网络频谱分配算法

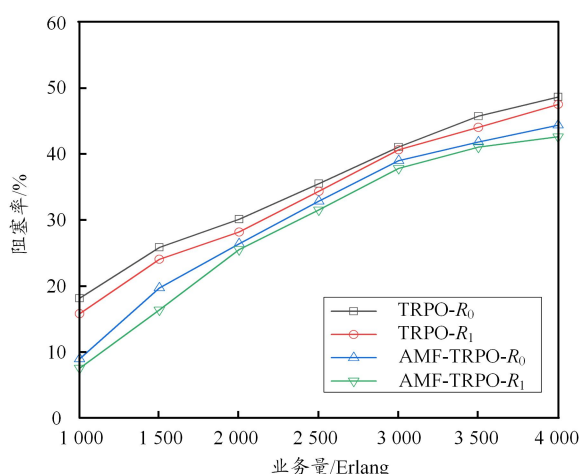


图3 4种算法阻塞率对比图

时, 相比于 TRPO- R_0 , TRPO- R_1 的阻塞率最高降低了约为2.3%, TRPO- R_1 在 C+L 波段下的阻塞率性能提升并不显著, 原因是可供业务分配的波段数量较少。若进一步将波段扩展至 S 或 U 波段, 该方法能够进一步降低网络阻塞率。然而, 所提算法在网络拓扑中的阻塞率均优于基准算法, 且所提 AMF-TRPO- R_1 算法在仿真的业务量范围内较基准 TRPO- R_1 算法最高降低了8.3%。这是因为所提出的算法根据业务的实际 OSNR 状况选择适当的调制格式, 相比于考虑最坏情况, 该方法能够尽可能为后续业务保留频谱资源, 从而降低业务的阻塞率。

当业务量为 1 000~4 000 Erlang 时, 4 种算法的频谱利用率情况如图 4 所示。可以看出, 所提出的 AMF-TRPO- R_1 算法的频谱利用率相较于 TRPO- R_1 算法最高提升了 9.6%。这是因为所提算法在进行资源分配时, 会依据业务当前的 OSNR 状况来选择调制格式, 而不是基于一个极端的假设, 即所有信道均处于活动状

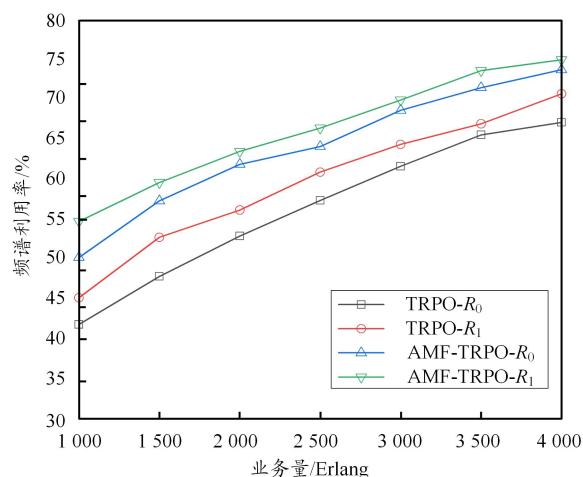


图4 4种算法频谱利用率对比图

态。因此, 在业务分配次数、业务量和其它参数配置相同的情况下, 本文算法可以承载更多的业务, 即占用更多的频隙, 从而实现更高的频谱资源利用率。

4 结束语

本文针对 C+L-EONs 的路由和频谱分配问题, 基于 DRL 中的 TRPO 算法, 提出了 AMF-TRPO 算法。该算法以 TRPO 为基础, 利用 DRL 智能化选择合适的调制格式, 根据业务当前实际所受物理层损伤状况进行决策, 同时还考虑未来业务对当前业务 QoT 的潜在影响。通过挖掘最佳路由路径并智能分配频谱资源来进一步优化 C+L-EONs 的性能。此外, 该算法还结合了 2 种奖励函数, 有效降低了网络的阻塞率并提高频谱利用率。仿真结果表明: 与基准的 TRPO 算法相比, 在仿真的网络负载范围内, 所提算法阻塞率最高降低 8.3%, 资源利用率最高提升了 9.6%。未来工作将致力于进一步提高系统性能, 包括增加更多训练拓扑和对训练超参数(如学习率、折扣因子等)进行敏感性分析、将强化学习算法与不同的启发式算法进行对比等, 从而更全面评估所提算法的性能。

参考文献:

- [1] SADEGHI R, CORREIA B, SOUZA A, et al. Transparent vs translucent multi-band optical networking: capacity and energy analyses [J]. Journal of Lightwave Technology, 2022, 40(11): 3486-3498.
- [2] MEHRABI M, BEYRANVAND H, EMADI M J. Multi-band elastic optical networks: inter-channel stimulated Raman scattering-aware routing, modulation level and spectrum assignment[J]. Journal of Lightwave Technology, 2021, 39(11): 3360-3370.
- [3] GUO N N, ZHANG K, LI Y C, et al. Impact of the band upgrade sequence on the capacity and capital expenditure of multi-band optical networks [J]. Journal of Optical Communications and Networking, 2023, 15(10): E1-E17.
- [4] JANA R K, CHATTERJEE B C, SINGH A P, et al. Quality-aware resource provisioning for multiband elastic optical networks: a deep learning-assisted approach[J]. Journal of Optical Communications and Networking, 2022, 14(11): 882-893.
- [5] ZHANG W C, YIN S, WANG Z H, et al. Routing, modulation level and spectrum assignment considering nonlinear interference in C+L+S-bands EONs[C]//IEEE. Proceedings of 2022 27th OptoElectronics and Communications Conference (OECC) and 2022 International Conference on Photonics in Switching and Computing(PSC). Toyama: IEEE, 2022: 1-3.
- [6] YAO Q Y, YANG H, BAO B W, et al. SNR re-verification-based routing, band, modulation, and spectrum assignment in hybrid c-c+ l optical networks[J]. Journal of Lightwave Technology, 2022, 40(11): 3456-3469.
- [7] 张晓雅, 高军诗, 顾仁涛, 等. 基于机器学习的光网络路由与频谱分配机制研究[J]. 电信工程技术与标准化, 2021, 34(4): 46-50.

- [8] CHEN X, LI B J, PROIETTI R, et al. DeepRMSA: a deep reinforcement learning framework for routing, modulation and spectrum assignment in elastic optical networks [J]. Journal of Lightwave Technology, 2019, 37(16): 4155–4163.
- [9] MORALES P, FRANCO P, LOZADA A, et al. Multi-band environments for optical reinforcement learning gym for resource allocation in elastic optical networks[C]//IEEE. Proceedings of 2021 International Conference on Optical Network Design and Modeling (ONDM). Gothenburg: IEEE, 2021: 1–6.
- [10] TERKI A B, PEDRO J, EIRA A, et al. Routing and spectrum assignment assisted by reinforcement learning in multi-band optical networks[C]//IEEE. Proceedings of 2022 European Conference on Optical Communication (ECOC). Basel: IEEE, 2022: 1–4.
- [11] SEMRAU D, KILLEY R I, BAYVEL P. A closed-form approximation of the Gaussian noise model in the presence of inter-channel stimulated Raman scattering[J]. Journal of Lightwave Technology, 2019, 37(9): 1924–1936.
- [12] GU Y, LI Y C, GUO N N, et al. L-band amplifier placement in C+L-band elastic optical networks[C]//IEEE. Proceedings of 2021 Asia Communications and Photonics Conference (ACP). Shanghai: IEEE, 2021: 1–3.
- [13] SHI C L, ZHU M, GU J H, et al. Deep-reinforced impairment-aware dynamic resource allocation in nonlinear elastic optical networks[C]//IEEE. Proceedings of 2021 Opto-Electronics and Communications Conference (OECC). Hong Kong: IEEE, 2021: 1–3.
- [14] NAKAGAWA M, KAWAHARA H, MASUMOTO K, et al. Performance evaluation of multi-band optical networks employing distance-adaptive resource allocation[C]//IEEE. Proceedings of 2020 Opto-Electronics and Communications Conference(OECC). Taipei: IEEE, 2020: 1–3.
- [15] TANG B, HUANG Y C, XUE Y, et al. Heuristic reward design for deep reinforcement learning-based routing, modulation and spectrum assignment of elastic optical networks [J]. IEEE Communications Letters, 2022, 26(11): 2675–2679.
- [16] GONZALEZ M, CONDON F, MORALES P, et al. Improving multi-band elastic optical networks performance using behavior induction on deep reinforcement learning[C]//IEEE. Proceedings of 2022 IEEE Latin-American Conference on Communications(LATINCOM). Rio de Janeiro: IEEE, 2022: 1–6.
- [17] 商盼盼, 任丹萍, 张娟, 等. 弹性光网络中面向业务精确迁移的频谱碎片整理算法[J]. 光通信技术, 2021, 45(10): 51–57.
- [18] BAO B W, YANG H, YAO Q Y, et al. Lorb: Link-oriented resource balancing scheme for hybrid c/c+ l band elastic optical networks[J]. Optical Fiber Technology, 2022, 74: 103071–1–1103071-10.