

基于强化学习的异构网络接入选择算法

张慧颖,马成宇,李月月,梁士达,盛美春

(吉林化工学院 信息与控制工程学院,吉林 吉林 132022)

摘要:针对异构网络接入选择时在提高吞吐量与保持高公平性方面所面临的挑战,提出了一种基于强化学习的异构网络接入选择算法,即近端策略优化(PPO)算法。该算法通过与环境进行交互并采样数据,以用户长期吞吐量和满意度最大化为优化目标,随机模拟用户位置,收集用户属性数据进行模型训练,以获取最优的网络接入点分配策略。仿真结果表明:与传统算法相比,PPO 算法在访问用户数达到最大值时,吞吐量可以提高 40%~70%,用户平均满意度仍可达 30%以上,用户公平指数可达 0.82。

关键词:强化学习;吞吐量;体验质量;公平指数

中图分类号:TN929 **文献标志码:**A **文章编号:**1002-5561(2024)04-0077-06

DOI:10.13921/j.cnki.issn1002-5561.2024.04.015

开放科学(资源服务)标识码(OSID):



Heterogeneous network access selection algorithm based on reinforcement learning

ZHANG Huiying, MA Chengyu, LI Yueyue, LIANG Shida, SHENG Meichun

(Jilin Institute of Chemical Technology Information and Control Engineering Institute, Jilin Jilin 132022, China)

Abstract: Aiming at the challenge of enhancing throughput and maintaining high fairness in heterogeneous network access selection, a proximal policy optimization (PPO) algorithm based on reinforcement learning is proposed. This algorithm interacts with the environment and samples data, aiming to maximize users' long-term throughput and satisfaction. It randomly simulates user locations, collects user attribute data for model training, and acquires the optimal network access point allocation strategy. The simulation results show that compared with traditional algorithms, when the number of access users reaches the maximum, the PPO algorithm can increase throughput by 40% to 70%, while the average user satisfaction can still exceed 30%, and the user fairness index can reach 0.82.

Key words: reinforcement learning, throughput, quality of experience, fairness indices

0 引言

近年来,随着光保真(LiFi)和无线保真(WiFi)异构网络的崛起,一种融合了 LiFi 高速传输与 WiFi 大面积覆盖优势的新型网络结构正逐渐成为未来室内无线通信领域备受瞩目的研究热点^[1]。然而,由于异构网络的特性,接入点(AP)的选择变得非常复杂^[2]。原因

在于,一方面,单个 LiFi AP 的覆盖区域相对有限,用户在移动过程中很容易超出其覆盖范围;另一方面, LiFi 与 WiFi 的覆盖区域存在重叠,这使得简单的切换策略难以实现;此外, WiFi 的大面积覆盖也容易导致异构网络过载。因此,如何选择合适的用户 AP 成为亟待解决的问题。

针对这一问题,众多学者提出了多种解决方案,大致可分为优化、模糊逻辑和机器学习等几类。文献[3]提出了一种动态负载平衡(LB)方案,旨在提升系统吞吐量和公平性。该方案通过综合考虑用户移动性和交换信令开销,对异构 LiFi 和 WiFi 网络中的负载均衡进行了深入研究。尽管该方案在吞吐量和用户公平性方面有所考虑,但其在复杂室内环境中的适应性仍有待提升。文献[4]提出了一种两阶段接入方法,用于异

收稿日期:2024-02-23。

基金项目:吉林省科技厅自然科学基金联合基金(N0.YDZJ202101ZYTS189)资助。

作者简介:张慧颖(1982—),女,吉林四平人,工学博士,副教授,硕士生导师,主要研究方向为空间光通信调制接收技术、可见光通信与定位技术。负责完成省部级项目 1 项,获省部级科学技术奖 1 项。



张慧颖,马成宇,李月月,等: 基于强化学习的异构网络接入选择算法

构 LiFi 和 WiFi 网络。设计了一个模糊逻辑系统来识别哪些用户应当连接至 WiFi; 在第二阶段, 则将剩余用户分配至 LiFi 网络。这种方法通过引入模糊逻辑, 有效地降低了决策复杂性并提升了网络吞吐量, 但忽略了用户满意度的考量, 因此并非最佳解决方案。文献[5]提出了一种基于 Q-Learning 的算法, 该算法通过设置最大化总吞吐量为奖励, 并结合实际距离与阈值距离的比率来确定用户位置, 从而使用户有可能接入离其最近的 AP。然而, 这种算法可能导致 AP 过载, 严重降低了用户的服务体验质量和用户之间的公平性。

基于以上分析, 本文提出一种基于强化学习的异构网络接入选择算法——近端策略优化(PPO)算法。

1 LiFi 和 WiFi 异构网络模型

室内 LiFi 和 WiFi 异构网络的示意图如图 1 所示。本文利用 LiFi 和 WiFi 网络实现室内下行通信, 因此在房间内安装了 4 组 LED 灯, 每组均可作为 LiFi AP 提供覆盖, 同时在房间中央设置了一个 WiFi AP, 以覆盖整个房间^[8]。

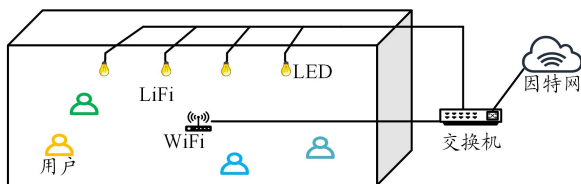


图 1 室内异构网络示意图

2 PPO 算法

PPO 算法的参数定义如下:

代理: 交换机被视作一个代理, 负责分配网络 AP。本文的策略旨在最大化长期吞吐量和用户满意度, 环境设定为包含 4 个 LiFi AP 和 1 个 WiFi AP 的场景。

奖励: 奖励函数 R 定义为 $R=aS+t$, 其中, a 为比例因子, S 为用户满意度, t 为吞吐量。由于用户满意度的取值为 0~1, 而吞吐量为 30~200 Mb/s, 两者间的数值差异显著。为了平衡这 2 个指标的权重, 本文引入了比例因子 a 来增加满意度的影响力。通过这一奖励机制, 可以为每个用户选择最佳的网络接入方式。

动作: 动作 $A_1 \sim A_5$ 均涉及从 5 个可用的 AP 中选择一个 AP 进行访问。每个用户在同一时间只能选择一个 AP 进行连接, 而不同用户可以选择不同的 AP。随着用户选择的 AP 不同, 网络环境也会发生相应变化。奖励会随着用户的访问过程不断累积, 直至所有用户获得最大的累积奖励。具体的操作定义如表 1 所示。

表 1 操作定义

动作	WiFi AP	LiFi AP ₁	LiFi AP ₂	LiFi AP ₃	LiFi AP ₄
A_1	1	0	0	0	0
A_2	0	1	0	0	0
A_3	0	0	1	0	0
A_4	0	0	0	1	0
A_5	0	0	0	0	1

注: 表中的 1、0 分别表示有、无访问。

PPO 算法流程图如图 2 所示。首先, 输入参数, 包括用户数量、AP 数量等。接着, 进行模型训练, 收集用户的当前状态, 并实时更新网络状态以反映用户接入的影响。此过程将持续进行, 直至所有用户均成功接入到对应的网络中, 并返回当前的奖励值。完成一轮训练后, PPO 算法会更新 Actor 网络和 Critic 网络的参数。其中, Actor 网络会沿增加优势函数的方向进行参数更新, 而 Critic 网络则会在接近真实奖励的方向上进行参数调整。通过多轮训练, 最终确定最适合的网络接入方案。

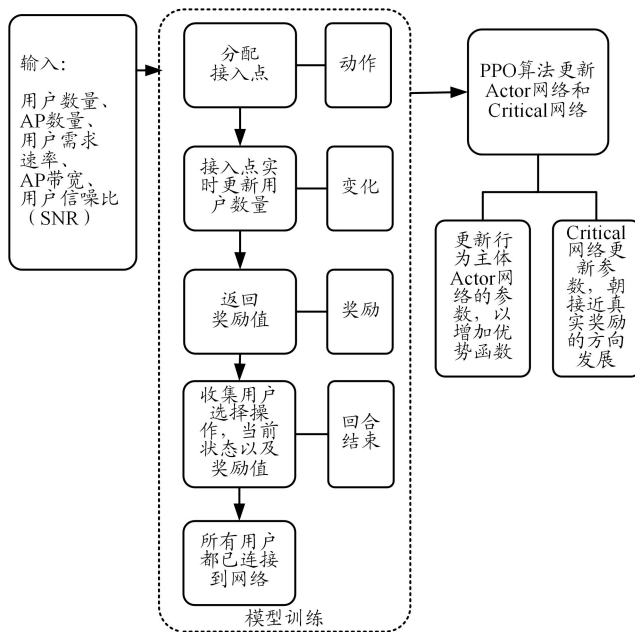


图 2 PPO 算法流程图

3 仿真结果与讨论

为了更加贴近现实, 本文用 Python 仿真模拟一个 10 m×10 m×3 m 的房间, WiFi AP 位于房间正中央, 覆盖整个房屋; 4 个 LiFi AP 均匀分布, 覆盖半径为 2.5 m^[9], 具体仿真参数如表 2 所示。

在最初的用户访问阶段, 用户 $U_1 \sim U_n$ 被随机分配

表2 信道参数

信道参数	数值
用户和AP之间的高度 h/m	2.5
光电探测器(PD)面积 s/cm^2	1
光学滤波器的增益 $T(\psi)$	1
半强辐射角 $\varphi_{1/2}/(^{\circ})$	70
PD的视场角 $\psi_c/(^{\circ})$	60
反射系数 ρ	0.8
光电转换效率 $c(\text{A/W})$	0.54
LiFi 传输光学功率 P_{opt}/W	20
LiFi 的带宽 $B_{\text{LiFi}}/\text{MHz}$	40
LiFi 的功率谱密度 $N_{\text{LiFi}}/(\text{A}^2/\text{Hz})$	1×10^{-21}
断点距离 d_{br}/cm	5
阴影衰落偏差 X/dB	5
中央载波频率 f_c/GHz	2.4
发射功率 $P_{\text{WiFi}}/\text{dBm}$	20
WiFi 信道的带宽 $B_{\text{WiFi}}/\text{MHz}$	20
WiFi 的功率谱密度 $N_{\text{WiFi}}/(\text{dBm}/\text{Hz})$	-174

到模拟房间的各个位置,并随后收集其接收信号强度(RSS)、SNR以及信号干扰加噪声比(SINR)等值。收集的数据集如表3所示。

表3 数据集

用户	LiFi	LiFi	LiFi	WiFi	WiFi
	RSS/lux	SNR/dB	SINR/dB	RSS/dBm	SNR/dB
U_1	811	55	7.2	-43	42
U_2	897	62	7.6	-50	33.4
...	...	...	...	...	...
U_n	359	40	0	-54	31.6

3.1 复杂度比较

本文采用了3种算法来计算异构网络的复杂度:

1)RSS算法。该算法为用户匹配SNR最高的AP^[20],其复杂度为 $O(N_{\text{AP}}N_U)$,其中 N_{AP} 表示AP的总数, N_U 表示访问用户的总数。

2)穷举搜索(ENUM)算法。该算法会搜索用户和AP之间的所有可能连接,它通过系统地枚举给定目标函数的所有可能动作并检查哪个特定动作最能满足目标函数^[21],其复杂性为 $O(N_{\text{AP}}N_U)$ 。

3)PPO算法。该算法的复杂度则取决于观察空间和动作空间的维数,复杂度为 $O(N_{\text{AP}}N_U^2+N_U^2+N_{\text{AP}}N_U)$ 。

3种算法的计算复杂度如图3所示。可以看出,与其它2种算法相比,本文提出的PPO算法的复杂度适中,虽然相较于最简单的RSS算法复杂度略有提升,但与最复杂的ENUM算法相比,在用户数量增加时,PPO算法的复杂度降低了许多。

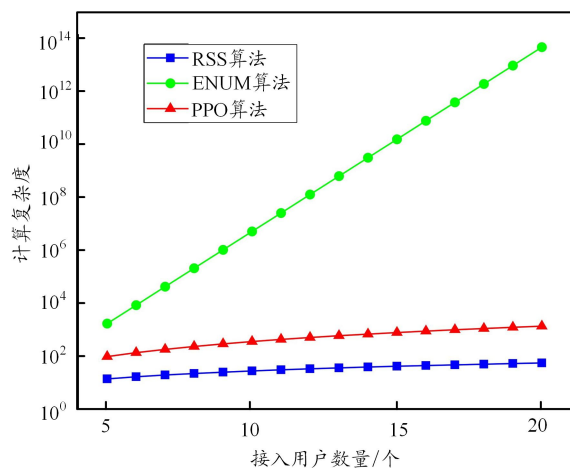


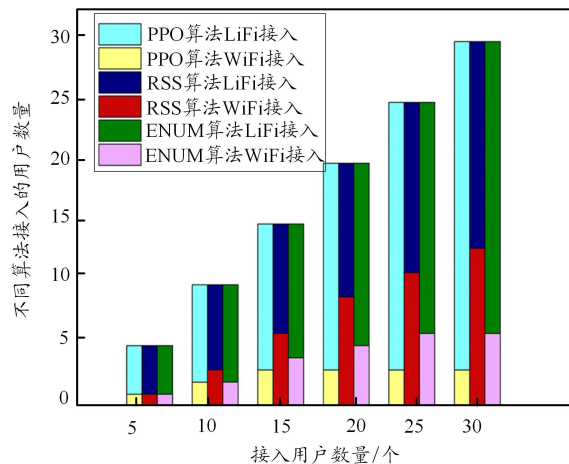
图3 3种算法的计算复杂度比较

3.2 访问的用户数量比较

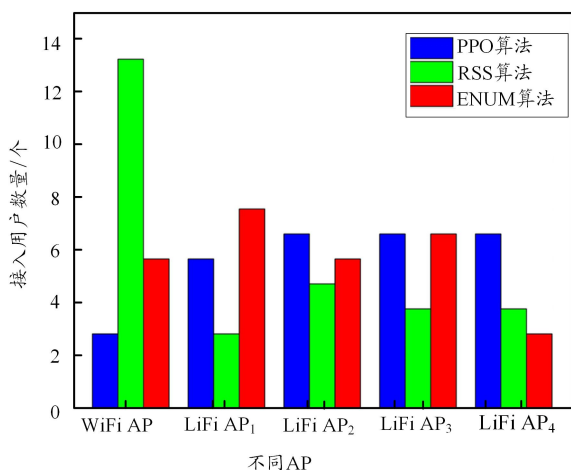
3种算法情况下访问用户数量比较如图4所示。从图4(a)可以看出,在LiFi AP和WiFi AP之间,只有PPO算法实现了相对均衡的WiFi用户分布。从图4(b)可以看出,在接入总用户数为30时,5个AP接入用户数量均为6个左右,从而有效避免了某个AP出现过载的情况。

3.3 可实现速率比较

3种算法的可实现速率对比情况如图5所示。从图5(a)可以看出,PPO算法的接入用户数量达到最大值时,平均可实现速率仍然略高于其它方法;从图5(b)可以看出,当接入用户数量达到30时,3种算法可实现的最高速率均为40 Mb/s左右,而PPO算法可实现的最

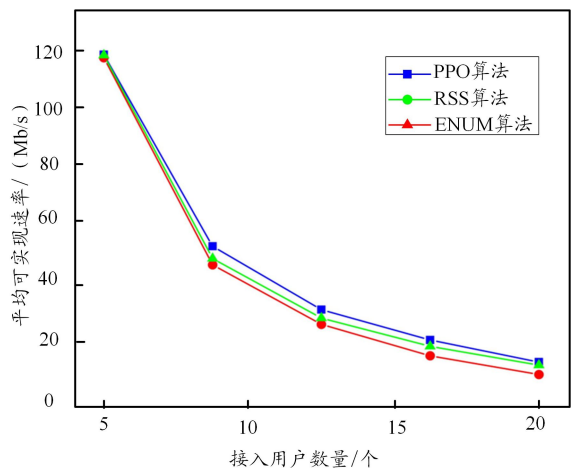


(a) 接入用户数量

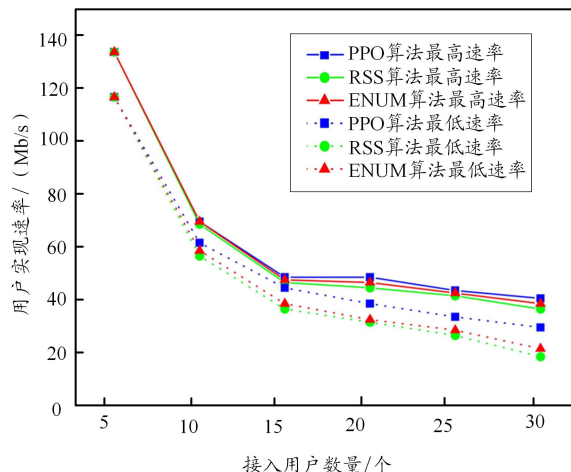


(b) 访问用户数量为 30 时接入用户数量

图 4 3 种算法情况下接入用户数量比较



(a) 平均可实现速率



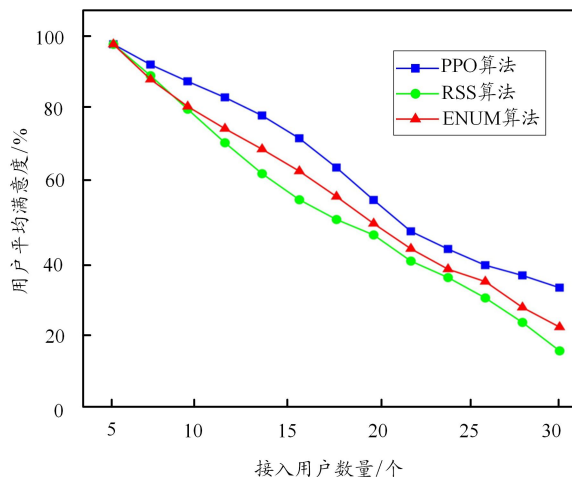
(b) 可以实现最高和最低速率

图 5 3 种算法的可实现速率比较

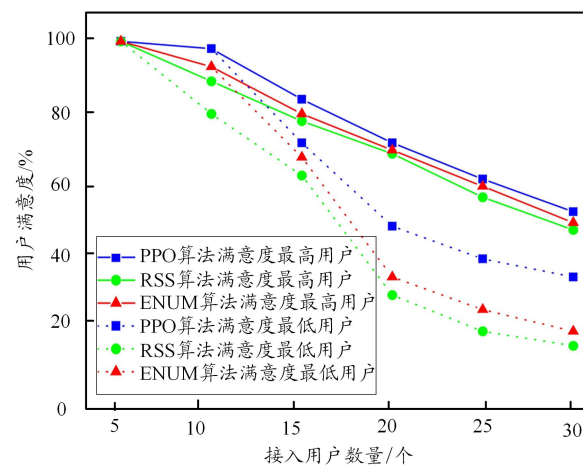
低速率为 30 Mb/s, 另外 2 种算法可实现的最低速率仅为 22 Mb/s 左右。这表明 PPO 算法能保证用户的可实现速率较高。

3.4 满意度比较

3 种算法的满意度比较如图 6 所示。从图 6(a)可以看出,PPO 算法使用户平均满意度达到 30%以上,显著优于其它算法;从图 6(b)可以看出,当接入用户数量达到 30 时,3 种算法可实现的最高满意度均为 60%左右,而 PPO 算法可实现的最低满意度为 40%,另外 2 种算法可实现的最低满意度仅为 20%左右。这表明 PPO 算法在提升用户满意度方面具有显著优势,能够确保所有用户获得较高的满意度水平,从而有助于提高用户的整体满意度。



(a) 用户平均满意度

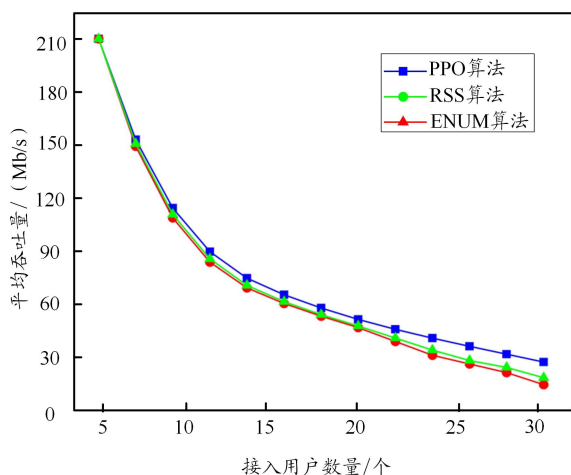


(b) 最高满意度比较和最差满意度

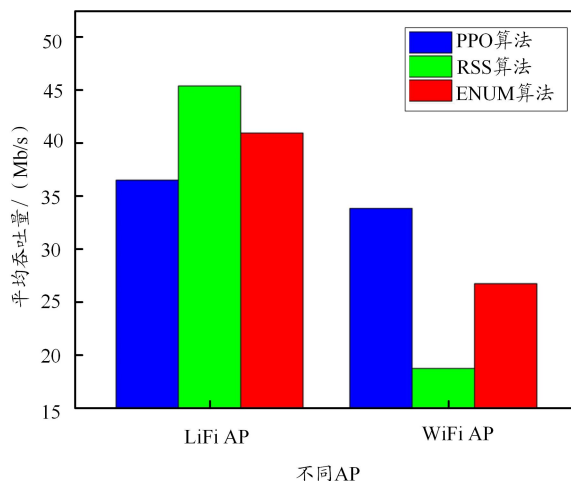
图 6 3 种算法的满意度比较

3.5 吞吐量比较

3 种算法的平均吞吐量对比情况如图 7 所示。从图 7(a)可以看出,当网络接入用户数量较少时,各算法的平均吞吐量差异并不显著。然而,当用户数量增加至 30 时,PPO 算法的表现尤为突出,与 RSS 算法、EBUN 算法相比,其平均吞吐量分别提升了 72%、



(a) 不同接入用户数量时的平均吞吐量



(b) 用户为 30 时实现的平均吞吐量

图 7 3 种算法可实现的平均吞吐量比较

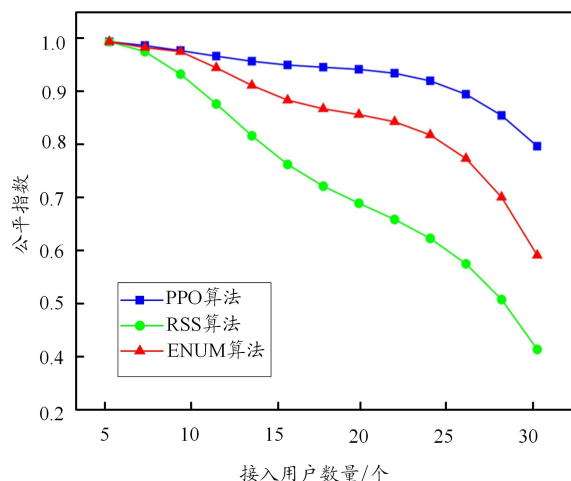


图 8 3 种算法的公平指数比较

40%。由此可见,在用户数量增多的情况下,PPO 算法能够显著提升平均吞吐量。从图 7(b)可以看出,当用户数达到 30 时,PPO 算法使得不同用户在访问 LiFi AP 和 WiFi AP 时的平均吞吐量保持大致平衡。相比

之下,其它算法往往只关注提升用户访问 LiFi 的吞吐量,而忽视了 WiFi 的吞吐量,这种不均衡会导致用户间的体验差异,进而对用户的满意度和网络负载均衡造成影响。

3.6 公平指数比较

3 种算法的公平指数比较情况如图 8 所示。可以看出,当接入用户数达到最大值时,PPO 算法的公平指数高达 0.82,而其它算法的公平指数分别为 0.47 和 0.63。这一结果充分证明了本文提出的 PPO 算法在解决户间公平性问题上的有效性。

4 结束语

本文提出了一种 PPO 算法。该算法旨在在满足最大吞吐量的同时,最大化用户满意度,为解决异构网络接入点的选择问题提供了新的解决思路。仿真结果表明,与 RSS 算法、ENUM 算法相比,PPO 算法具有更小的复杂度,在吞吐量和满意度方面均表现优异。特别是在访问用户数达到最大值时,吞吐量能提升 40%~70%,且平均满意度依然保持在 30% 以上。此外,用户之间的公平指数更是高达 0.82。未来,我们的目标是进一步优化该算法,使其更好地适应复杂环境,并提升其性能和收敛速度。

参考文献:

- [1] WANG Y, HAAS H. Dynamic load balancing with handover in hybrid Li-Fi and Wi-Fi networks[J]. Journal of Lightwave Technology, 2015, 33(22): 4671–4682.
- [2] AHMED A, BOULAHIA L M, GAITI D. Enabling vertical handover decisions in heterogeneous wireless networks: a state-of-the-art and a classification[J]. IEEE Communications Surveys & Tutorials, 2013, 16(2): 776–811.
- [3] WU X, SAFARI M, HAAS H. Access point selection for hybrid Li-Fi and Wi-Fi networks [J]. IEEE Transactions on Communications, 2017, 65(12): 5375–5385.
- [4] ALENEZI A M, HAMDI K A. Reinforcement learning approach for hybrid Wi-Fi-VLC networks[C]//IEEE. Proceedings of 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring). Antwerp: IEEE, 2020: 1–5.
- [5] WU X, SOLTANI M D, ZHOU L, et al. Hybrid LiFi and WiFi networks: a survey[J]. IEEE Communications Surveys & Tutorials, 2021, 23(2): 1398–1420.
- [6] MA Y, ZHOU G, WANG S. WiFi sensing with channel state information: a survey[J]. ACM Computing Surveys (CSUR), 2019, 52(3): 1–36.
- [7] HENRY P S, LUO H. WiFi: what's next?[J]. IEEE Communications Magazine, 2002, 40(12): 66–72.
- [8] HAAS H. Visible light communication [C]//IEEE. Proceedings of 2015 Optical Fiber Communications Conference and Exhibition(OFC). Los Angeles: IEEE, 2015: 1–72.

张慧颖,马成宇,李月月,等: 基于强化学习的异构网络接入选择算法

- [9] HAAS H, YIN L, WANG Y, et al. What is LiFi?[J]. Journal of light-wave technology, 2015, 34(6): 1533–1544.
- [10] HAAS H. LiFi is a paradigm-shifting 5G technology[J]. Reviews in Physics, 2018, 3: 26–31.
- [11] SMITH P J, DMOCHOWSKI P A, SHAFI M, et al. On the computation of per-user rate distributions from sum-rate distributions[J]. IEEE Wireless Communications Letters, 2014, 3(4): 381–384.
- [12] JAIN R K, CHIU D M W, HAWKINS W R. A quantitative measure of fairness and discrimination[EB/OL]. [2024–02–23]. <https://www.xueshufan.com/publication/1630733292>.
- [13] JAIN R, DURRESI A, BABIC G. Throughput fairness index: an explanation[EB/OL]. [2024–02–23]. https://www.cse.wustl.edu/~jain/atmf/ftp/af_fair.pdf.
- [14] XIAO Y, ROSDAHL J. Throughput and delay limits of IEEE 802.11[J]. IEEE Communications letters, 2002, 6(8): 355–357.
- [15] Kaelbling L P, Littman M L, Moore A W. Reinforcement learning: a survey[J]. Journal of Artificial Intelligence Research, 1996, 4: 237–285.
- [16] ABEBE S G, GEBEYEHU Z. Reinforcement learning approach handover management for hybrid LiFi and WiFi networks[EB/OL]. [2024–02–23]. <https://10.2139/ssrn.4168600>.
- [17] HAAS H, YIN L, CHEN C, et al. Introduction to indoor networking concepts and challenges in LiFi[J]. Journal of Optical Communications and Networking, 2020, 12(2): A190–1–A190–14.
- [18] ZHOU T Q, HUANG Y M, YANG L X. Joint user association and resource partitioning with QoS support for heterogeneous cellular networks[J]. Wireless Personal Communications, 2015, 83(1): 383–397.