

华中科技大学

本科毕业设计[论文]

基于深度强化学习的 通感一体化网络资源分配优化算法研究

院 系 电子信息与通信学院

电子信息工程（基于项目信息类专业
教育实验班（2+2））201901 班

姓 名 董浣羽

学 号 U201815671

指导教师 黑晓军

2023 年 5 月 25 日

学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的
研究成果。除了文中特别加以标注引用的内容外，本论文不包括任何其他个人或集
体已经发表或撰写的成果作品。本人完全意识到本声明的法律后果由本人承担。

作者签名：年 月 日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保障、使用学位论文的规定，同意学校保留
并向有关学位论文管理部门或机构送交论文的复印件和电子版，允许论文被查阅
和借阅。本人授权省级优秀学士论文评选机构将本学位论文的全部或部分内容编
入有关数据进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论
文。

本学位论文属于 1、保密 ☐ ，在 年解密后适用本授权书。

2、不保密 ☐ 。

（请在以上相应方框内打“√”）

作者签名：年 月 日

导师签名：年 月 日

摘 要

通感一体化,将通信系统与无线感知系统集成为一体,使得无线网络具备对物理世界的感知能力,如目标定位与追踪、环境成像、环境重建等。通感一体化系统,被认为是下一代通信技术在诸多应用领域的关键推动因素,包括但不限于智能车联网、智能家居和智慧城市等领域。

资源调度是影响通感一体化系统中通信与感知的效率与质量的关键。本文首先针对现有多用户雷达感知与下行通信共存的场景下,使用时分复用提升系统效率的功率资源分配的论文进行算法复现及性能评测,对其通信吞吐量、感知准确率、有效吞吐量进行分析评估,分析其优势和劣势提出改进措施。

基于以上复现和分析结果,本文设计了一个多用户、多输入多输出的通感一体化网络资源调度系统,采用雷达感知与下行通信相结合,使用空分复用提高通信系统效率和性能,并对场景、信号、信道进行建模,构建了完整的数字信号处理流程和相应评价指标。本文进一步将上述系统的资源分配问题建模为一个多用户空分复用场景下的功率分配问题,提出了基于深度强化学习算法 PPO 的多用户空分复用的资源调度系统设计方案,利用 Python 实现仿真,并设计实验验证其通感联合速率、通信和速率及雷达信息估计率和的指标性能。相比基于均匀功率分配的基线算法,本文提出的调度算法具有智能化、可调控性能目标的特点。在三种不同的用户数配置下的仿真实验中,其通感联合性能表现优于基线 5%~7%。

关键词: 通感一体化; 强化学习; 资源调度; 功率分配

Abstract

Integrated sensing and communication (ISAC) networks enable the wireless communication network to sense the physical world, such as target positioning and tracking, environmental imaging and reconstruction, etc. The ISAC network is considered to be a key enabler of many future areas of next-generation communications, including but not limited to smart vehicle networks, smart homes and smart cities.

Resource scheduling is crucial to the efficiency and quality of communication and sensing in an integrated communication and sensing system. The purpose of this thesis is to study the resource scheduling problem in ISAC systems. Firstly, this paper replicates and evaluates existing work for time resource scheduling in a multi-user time-division multiplexing ISAC system, and analyzes its communication throughput, sensing accuracy and effective throughput. This thesis further evaluates its advantages and disadvantages, and proposes improvement directions.

In order to study the resource scheduling problem in ISAC system, this thesis designs a multi-user, multi-input and multi-output ISAC network resource scheduling system, using radar sensing combined with downlink communication. Spatial division multiplexing is adopted to improve the efficiency and performance of the system,

In this thesis, we construct a power allocation problem in a multi-user spatial division multiplexing scenario based on proximal policy optimization (PPO), and propose a reasonable resource scheduling system design scheme. Simulation experiments are conducted to evaluate the joint ISAC rate, communication sum-rate and radar information estimation sum-rate. The joint performance of sensing and communication of the proposed algorithm outperforms the fully-equal power allocation baseline algorithm by 5%~7% under three different scenario configurations.

Key Words: integrated sensing and communication; reinforcement learning; resource allocation; power allocation

目 录

摘 要	I
Abstract	II
1 绪论	1
1.1 课题背景与意义	1
1.2 国内外研究现状以及发展趋势	2
1.3 课题目标	2
1.4 主要内容及创新点	3
2 相关技术概要	4
2.1 通感一体化网络介绍	4
2.2 强化学习介绍	5
2.3 本章小结	6
3 多用户时分复用通感一体化网络资源调度算法复现.....	7
3.1 系统模型	7
3.2 问题建模	8
3.3 问题求解	9
3.4 实验复现结果及分析	12
3.5 存在的问题	14
3.6 本章小结	14
4 多用户空分复用的通感一体化网络系统建模.....	15
4.1 场景介绍	15
4.2 下行通信信号模型	16
4.3 感知信号模型	16
4.4 多用户通信信号检测模型	17
4.5 通信性能指标	18
4.6 多目标雷达信号检测模型	18
4.7 感知性能指标	19
4.8 本章小结	20
5 基于 PPO 的多用户空分复用场景下的资源调度算法设计与实现.....	21
5.1 问题建模	21
5.2 马尔可夫决策过程设计	23
5.3 算法流程	25
5.4 本章小结	27
6 仿真与性能评测	28
6.1 仿真场景设置	28
6.2 算法训练	30
6.3 通感联合性能	32
6.4 通信性能	33

6.5	感知性能	33
6.6	本章小结	34
7	总结与展望	35
7.1	工作总结	35
7.2	未来展望	36
致谢		37
参考文献		38

缩写词汇表

英文缩写	英文全称	中文翻译对照
AI	Artificial Intelligence	人工智能
CACC	Cooperative Adaptive Cruise Control	合作自适应巡航控制
DFRC	Dual-Functional Radar-Communication	双功能雷达通信
DNN	Deep Neural Network	深度神经网络
DoA	Direction of Arrival	波达方向
DRL	Deep Reinforcement Learning	深度强化学习
GAE	General Advantage Estimation	广义优势估计
GNSS	Global Navigation Satellite System	全球导航卫星系统
IoT	Internet of Things	物联网
ISAC	Integrated Sensing And Communication	通感一体化
KPI	Key Performance Indicator	关键绩效指标
MDP	Markov Decision Process	马尔可夫决策过程
MIMO	Multiple Input Multiple Output	多输入多输出
MU	Multiple User	多用户
MVDR	Minimum Variance Distortionless Response	最小方差无失真响应
NR	New Radio	新无线电
PPO	Proximal Policy Optimization	近端策略优化
RSU	Road Side Unit	路侧单元
SISO	Single Input Single Output	单输入单输出
TRPO	Trust Region Policy Optimization	信赖域策略优化
UAV	Unmanned Aerial Vehicle	无人驾驶飞行器
V2V	Vehicle to Vehicle	车际互联
ZF	Zero Forcing	迫零

1 绪论

1.1 课题背景与意义

随着通信行业的发展,无线通信的服务量空前激增。“万物互联”是 5G 时代的一大愿景,随着 B5G/6G 时代的到来,通信服务提供商不断在探索新的频谱资源,信号波段不断提高,逐步向着毫米波、太赫兹 (THz) 网络发展^[1]。最初,为充分利用频带资源,将仅限于雷达系统的频段供通信业务共享使用,相关研究人员提出了一种新型的网络架构——通感一体化网络 (Integrated Sensing and Communication, ISAC),探究雷达和通信频谱共享、雷达-通信合作、双功能雷达通信系统 (Dual-Functional Radar-Communication, DFRC) 的可行性^[2]。近年,通感一体化的概念与范畴在^[3,4]中被正式定义,吸引了工业界和学术界的广泛研究关注。与单独的感知或通信系统相比,通感一体化网络的设计方法由两个显著优势。首先,频谱、功率和硬件平台共享,使感知和通信的效率得到提高。其次,感知与通信业务之间的可以互相为对方提供支持,因此通感一体化系统可以进一步获得联合的性能提升^[2]。因此,通感一体化网络被认为是下一代通信的诸多应用领域的关键推动因素,包括但不限于智能车联网、智能家居和智慧城市。

感知,将成为下一代无线通信网络的内生能力。下一代无线通信将使用更高的频段(从毫米波到太赫兹)、具有更高的网络带宽,并实现大规模天线阵列的密集应用^[1]。下一代无线通信的三个特点——更高的频段、更高的带宽、更密集的大规模天线阵列应用,将使得通信系统内部,能够进一步集成感知功能,称为通感一体化系统。通信系统作为一种广泛分布的、可以发射电磁波的实体系统,其发射得电磁波得反射和散射信号,将包含物理世界得实际信息。又因原生无线通信系统具有高覆盖的特点,向无线通信系统集成感知功能,使得感知系统也无处不在。这种感知系统已被实现并被证明具有高精度的定位、高精度成像甚至环境重建功能^[5,6]。一方面,通感一体化系统种,感知系统反哺于通信系统,可提高通信性能——例如,感知提供的用户位置信息,可以实现更准确的波束对准,更敏捷的波束失败恢复、更低的信道追踪开销^[7]。另一方面,通感一体化系统可实现对物理世界的观测与重建,可以用于建立物理世界与虚拟世界的联系。因此,通感一体化技术,对于实现“数字孪生”、“元宇宙”,是一项至关重要的技术^[8]。

通感一体化网络架构的引入,带来一系列新的问题。数字信号处理流程的变化,系统目标不再局限于通信性能,使得传统无线网络中的资源管理与调度方式不能直接用于通感一体化的网络。本课题聚焦于通感一体化网络中的资源调度问题,关注通信业务与感知业务共存的无线通信系统中,如何构建有实际意义的通感一体化场景建模,设计合理的资源调度算法,保证通信业务与感知业务的性能。

1.2 国内外研究现状以及发展趋势

根据传感目标是否能发送传感信号,通感一体化一般可分为两种类型,即基于设备的通感一体化,和无设备的通感一体化^[9]。对于基于设备的通感一体化,感知目标配备了用于发送感知信号的设备。然而,基于设备的通感一体化有着更高的硬件开销,同时,感知信号的传输会造成相对较长的时间延迟。

许多研究工作都致力于利用无设备的通感一体化来改善传感和通信性能。在^[10]中,作者定义了无设备通感一体化系统中的感知性能指标与通信性能指标,并提出了一个统一的功率与带宽资源分配框架。在^[11]中,作者提出了一种基于时分复用的多用户通感一体化资源调度策略,同时讨论了通信与感知业务相耦合的情况下的有效吞吐最大化问题。在^[12]中,作者创新提出了非正交多址接入在通感一体化系统中带来的性能提升。在^[13]中,作者提出了一种通感一体化的功率分配策略,在保证通信速率的同时,最大限度地提高检测准确性的概率。

然而,对于通感一体化系统中资源调度问题,现有研究仍有改进空间。^[10]中虽然提出了一个功率带宽分配策略的统一框架,但缺少对问题求解方式的具体探讨。^[11]中的工作只考虑了通信与感知业务耦合的情况,缺少普适性。^[12]中作者虽然将非正交多址接入应用于通感一体化系统,提出了相应的资源分配算法,但对实际的感知性能指标没有做相应讨论。此外,之前的工作鲜少关注强化学习在通感一体化资源调度问题上的应用。因此,本课题基于现有通感一体化系统资源调度算法,重点研究通感一体化系统中基于强化学习的用户调度与资源分配策略,以提高系统的感知与通信性能。

1.3 课题目标

针对本课题,将首先对参考文献^[11]的工作进行复现,对多用户下行通感一体

化场景进行建模、分析与仿真,进而研究当前建模存在的泛化性问题,并针对该问题,对场景建模进行扩展,在新场景中设计实现基于强化学习的资源调度算法,并设计实验,验证算法的性能。

- 首先,基于参考文献^[11]中提出的基于时分复用的下行多用户通感业务模型,分析通信与感知业务关键性能指标——吞吐量与检测率与资源调度方案的定量关系。进而利用 Python 对参考文献^[11]中的模型与算法进行仿真设计,验证模型的准确性。
- 基于参考文献^[11]中提出的模型,进而考虑感知与通信不耦合的情况,对模型进行重新设计,并基于重新设计的理论模型得出通信与感知性能指标与资源分配方案的关系。利用 Python 对重新设计的理论模型进行仿真设计。
- 基于重新设计的理论模型,设计基于强化学习的资源调度算法。利用 Python 进行仿真设计,验证其在通信性能、感知性能等方面的效果。

1.4 主要内容及创新点

本文根据上文提到的相关工作发展趋势,首先复现一种基于时分复用的多用户通感一体化系统,根据对该系统的分析评测,提出一种多用户、多输入多输出的通感一体化系统,采用多用户信号检测技术实现同时服务于多个通信用户,采用雷达波束赋形-信号检测的技术实现同时检测多个感知用户,从而提高了系统的实时性能容量;本文进而构建该系统中的功率资源调度问题,设计了基于强化学习的功率分配算法,验证其性能优于基线算法。

本课题系统的创新点在于:

- 现有大部分工作关注单感知用户的雷达感知,而本文利用雷达波束赋形-信号检测的技术实现同时检测多个感知用户,探究多用户、多输入多输出的通感一体化系统中的资源分配问题。
- 本文创新采用强化学习的方式,研究多用户、多输入多输出的通感一体化系统中的资源分配问题。

2 相关技术概要

本章节将进行相关技术的概要,着重介绍通感一体化网络的定义、发展情况与应用前景;同时,本文还将对强化学习做简要介绍,并阐述其在无线网络优化中的应用价值。

2.1 通感一体化网络介绍

在上一章节,本文对通感一体化网络(Integrated Sensing and Communication, ISAC)做了简单的介绍。在本小节,本文将进一步介绍通感一体化网络在多个领域的应用现状与前景。

增强目标定位和跟踪:定位一直是现有蜂窝网络(从1G到未来的6G)的标准化、落地和蜂窝小区构建的一个关键环节^[14]。由于低带宽和较少的天线数量,造成目前测距和测角度的分辨率低。目前大多数蜂窝网络(即4G和5G网络)只能提供米级精度的测量数据,以协助全球导航卫星系统(Global Navigation Satellite System, GNSS)。根据5G新无线电(New Radio, NR)第17版的关键参数指标(KPI)^[15],在工业物联网领域,最高要求水平/垂直定位精度达到0.2米/1米,但这无法满足未来应用的要求。特别是在室内环境中,准确定位用户位置的要求比室外更高,例如,室内人类活动识别^[16](~1厘米),自主机器人和制造^[17](~5毫米)。另一方面,目前基于蜂窝的定位技术大多以基于设备的方式实现,其中无线设备(如智能手机)通过与其他部署的无线设备(如Wi-Fi接入点或基站(BS))的信号交互以及几何位置关系来计算其具体位置。然而,基于设备的方法限制了对定位对象的选择,而且不能适用于不同的场景。通过发掘多径传播的有用信息与多普勒处理,与目前的定位技术相比,基于ISAC的蜂窝网络能够提高定位精度。除此之外,具有传感功能的蜂窝网络并不局限于用智能手机确定某个物体的位置,而且还非常适合从感知周边的物理环境。

区域成像:射频成像技术可以生成高分辨率、昼夜不分的图像,用于环境监测、气候变化研究和与安全相关的应用等多种用途^[18]。更重要的是,与基于摄像头的成像技术相比,它对环境的干扰性较小,可以在不暴露在周围环境中情况下,获取周围环境的信息。由于上一代蜂窝系统大多利用窄带技术,其范围分辨率大约为米级,不支持高分辨率的服务。近年来,随着毫米波段甚至更高频段的推广和大规模

多输入多输出技术的落地部署,未来的基站有可能通过合作感知和成像,达到更高的探测覆盖,更高的测距和测角分辨率。在这种情况下,综上所述,无线电接入网络可以作为一个分布式的毫米波雷达系统。因此,未来的蜂窝网络 and 用户设备(User Equipment, UE)可以探测的环境,这将进一步支持高层应用,如数字孪生、虚拟现实等^[7]。此外,由于更高的信号频率使得成像分辨率显著提高,未来的蜂窝网络还将支持基于空间/位置感知的服务。最后,具有感知成像能力的基站和用户设备可以为传统电信运营商提供额外的商业价值,作为一种新的计费服务。

智能车联网:在紧密间隔、计算机控制的自动驾驶车辆调度系统,使得高速公路容量的增加和乘客舒适度的提高。目前的自动驾驶车辆调度方案主要是通过传统的领导者-追随者框架,基于合作自适应巡航控制(Cooperative Adaptive Cruise Control, CACC)^[19,20]的实现,需要多跳车辆对车辆(Vehicle to Vehicle, V2V)通信,所有排队的车辆之间需要传递每辆车的状态信息。然而,多跳通信的高延迟也导致了排队车辆状态信息的不同步现象。当排队非常长且队列状态高度动态时,这个现象尤其严重。在这种情况下,不了解情况变化的排队车辆会增加控制风险。路侧单元(Road Side Unit, RSU)作为车辆基础设施,可提供一种更可靠的方法来管理车辆队列,因为它同时为周边多辆车服务^[21,22]。在RSU上配备的雷达感知功能,将提供一种快速和廉价的方式来获取大量车辆的状态^[23,24]。

2.2 强化学习介绍

强化学习是机器学习最重要的研究方向之一,在过去20年里对人工智能(Artificial Intelligence, AI)的发展产生了重大影响^[25]。强化学习使用智能体(Agent)与环境的反复交互过程中收到反馈,并不断更新智能体的策略。在这个过程中,智能体可以做出动作(Action),收集对环境的观测(Observation),然后自动调整策略以达到最佳策略^[26]。然而,在大规模问题中,朴素的强化学习需要大量的时间收敛到最佳策略,因为智能体必须通过探索,逐渐获得整个系统的知识。而大规模问题会使搜索空间急剧膨胀,因而朴素的强化学习不适用于大规模问题。深度学习作为一种新的突破性技术被引入,深度神经网络(Deep Neural Network, DNN)对高维数据、复杂系统的强大拟合能力,可以克服强化学习的局限性,从而形成一类新的算法,即深度强化学习(Deep Reinforcement Learning, DRL)^[25]。

下一代通信网络的典型应用场景，如物联网（Internet of Things, IoT）和无人驾驶飞行器（Unmanned Aerial Vehicle, UAV）网络，变得更加分散和自主。在这样的网络中，网络实体需要在本地做出决策，以便在网络环境不确定的情况下使网络性能最大化，如链路自适应、发射功率控制、接入选择等^[27]。网络通信环境随机高，其中很多决策问题，可以被建模为马尔科夫决策过程（Markov Decision Process, MDP）^[28]，进而采用各种算法来求解。动态编程，价值迭代，以及强化学习技术，可以被用来解决 MDP 问题。现代网络是有着较大的规模与复杂性，这些算法在现实场景下的计算复杂度有可能远超设备的本地计算能力。因此，深度强化学习被认为是为克服这一挑战的解决方案。

深度强化学习的深度神经网络具有高维拟合能力，经过合理设计，可以发掘数据之间的高维的、复杂的关系，建立高维映射。使得无线网络通信中的各种非凸和复杂的问题都可以考虑采用强化学习求解，如功率控制、用户调度、资源调度等问题^[29]。在没有完整和准确的网络信息的情况下，达到最优或接近最优的网络性能。深度强化学习允许网络实体感知通信和网络环境。因此，通过使用深度强化学习，网络实体，如用户可以在不知道信道模型和移动模式的情况下在诸如基站选择、信道选择、交接决策、缓存和任务卸载决策等问题场景中学习最佳策略。通过深度强化学习方法，网络实体可以采集环境数据，并在本地环境中离线学习最佳策略，而不一定需要在实际环境中完成模型的训练。这不仅减少了通信开销，而且还提高了网络的安全性和稳健性。深度强化学习在具有高维状态和动作空间的问题上有着速度优势。因此。在大规模的网络中，尤其是有成千上万设备的物联网系统，深度强化学习使得网络控制器或物联网网关可以根据网络环境进行动态的功率控制、用户调度、资源调度等，以达到更好的网络性能。

2.3 本章小结

本章介绍了本文问题场景下的系列背景。首先，本章进一步介绍了通感一体化网络在多个领域的应用现状与前景，阐述了通感一体化网络在未来无线网络中的应用优势。其次，本章对强化学习进行了简单介绍，并阐述了强化学习在无线网络中的应用优势与前景。

3 多用户时分复用通感一体化网络

资源调度算法复现

本章节将对一篇以时分复用方式构建的通感一体化资源调度系统的论文进行算法复现以及效果评测。本章将会对该资源调度系统的通信吞吐量、感知准确率、有效吞吐量进行分析评估。最后，本章会分析该资源调度系统的优势及劣势，并提出相应改进方向。在下一章节，本文会针对上述问题，设计新的通感一体化资源调度系统。

3.1 系统模型

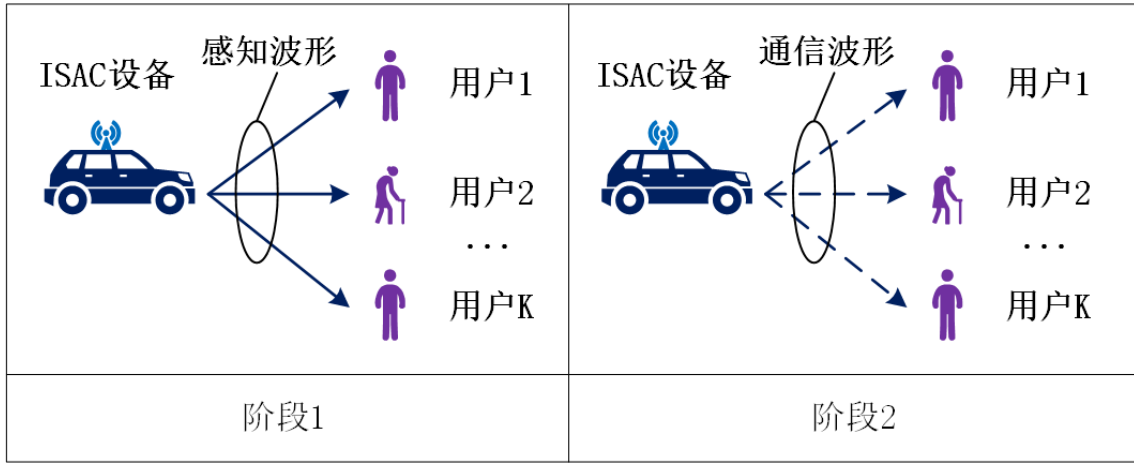


图 3-1 多用户时分复用的通感一体化系统场景示意图

如图所示，本系统中存在一个多功能机动车以及 K 个行人，文中使用 $\mathcal{K} = \{1, 2, \dots, K\}$ 来表示所有行人组成的集合。机动车装备有 ISAC 设备可以发射波形来感知行人的状态并向行人提供下行通信业务。用户之间、以及雷达感知与下行通信采用时分复用的方式。整个流程分为两阶段：（1）第一阶段，行人通过雷达回波信号被机动车检测到；同时，机动车通过检测算法识别行人的；（2）第二阶段，机动车向行人发射通信信号，进行下行通信。文中使用 c_k 表示分配给第 k 个行人的感知轮数，每轮感知时长为 T_0 ，那么对于第 k 个行人，感知时间为 $c_k T_0$ 。因此，对于所有用户而言，总的感知时间可以表达为：

$$T_s^{\text{ph-l}} = \sum_{k=1}^K c_k T_0. \quad (3-1)$$

根据^[11], 对于行人状态分类的准确率(下文统一称为“感知准确率”)可以与分配到的感知轮数有如下关系:

$$\Theta_k(c_k) = \gamma_k - \alpha_k c_k^{-\beta_k}. \quad (3-2)$$

其中, Θ_k 代表感知准确率, γ_k 、 α_k 以及 β_k 均为参数。可以从上式得到, 分配给一个行人的感知轮数越多, 这个行人将会得到更高的感知准确率; 同时, 这种感知准确率的增幅有一个上限。

在第二个阶段, 机动车以时分复用的方式向行人发送信息。采用 t_k 表示第 k 个用户的下行通信时间。第二阶段总体的时间消耗为:

$$T_c^{\text{ph-II}} = \sum_{k=1}^K t_k. \quad (3-3)$$

那么, 感知与通信两个阶段的总体时间消耗为:

$$T^{\text{tot}} = T_s^{\text{ph-I}} + T_c^{\text{ph-II}}. \quad (3-4)$$

根据香农信道容量^[30], 机动车对行人 k 的下行可达发送数据量可以表示为可达速率与时间的乘积:

$$R_k^{\text{data}} = \frac{t_k}{T^{\text{max}}} B \log_2 \left(1 + \frac{g_k P^{\text{com}}}{\delta} \right). \quad (3-5)$$

作者假设了实际下行通信的有效性与检测相关, 定义了有效通信吞吐量:

$$R_k^{\text{eff}} = \Theta_k(c_k) \frac{t_k}{T^{\text{max}}} B \log_2 \left(1 + \frac{g_k P^{\text{com}}}{\delta} \right). \quad (3-6)$$

公式(3-5)与公式(3-6)中的 B 表示通信带宽, g_k 表示信道增益, P^{com} 表示通信发射功率, δ 表示噪声功率。

3.2 问题建模

文中构建了一个最大化有效吞吐量和的优化问题。其中, 有效吞吐量和表示为各个行人的吞吐量对数和。作者采用这个目标函数的目的是使得各个行人的有效吞吐量达到比例公平:

$$R^{\text{sum-eff}} = \max \sum_{k=1}^K \ln(R_k^{\text{eff}}). \quad (3-7)$$

作者构建了一个最大化有效吞吐量和的优化问题, 并且对时间开销、能量开销以及每个行人的通信时间做了上限约束:

$$\text{maximize } \sum_{k=1}^K \ln(R_k^{\text{eff}}), \quad (3-8)$$

$$\text{subject to: } \sum_{k=1}^K T_0 c_k + \sum_{k=1}^K t_k \leq T^{\max}, \quad (3-9)$$

$$\sum_{k=1}^K T_0 c_k P^{\text{sen}} + \sum_{k=1}^K t_k P^{\text{com}} \leq E^{\max}, \quad (3-10)$$

$$0 \leq t_k \leq T^{\max}, \quad (3-11)$$

$$\text{variables: } c_k \in \{0, 1, \dots, C\}, t_k, \forall k \in \mathcal{K}. \quad (3-12)$$

3.3 问题求解

问题(3-8)是一个非凸问题,因为每个用户的感应周期 c_k 是离散的。为了求解这个非凸问题,文中将原始问题分解成两个子问题,并以分层方式求解。本节首先对问题(3-8)进行分解,然后针对顶层问题,提出一种基于短视分配算法(Myopic Allocation Algorithm)的求解方案。

问题(3-8)的重要特征如下:给定变量 $\{c_k, \forall k \in \mathcal{K}\}$ 的值,问题(3-8)的目标函数对于 $\{t_k, \forall k \in \mathcal{K}\}$ 来说是严格凸的。因此,根据这一特征,文中提出问题的分解方案。

在给定 $\{c_k, \forall k \in \mathcal{K}\}$,优化 $\{t_k, \forall k \in \mathcal{K}\}$ 。在给定 $\{c_k, \forall k \in \mathcal{K}\}$ 的情况下,约束(3-9)以及(3-10)均为关于 $\{t_k, \forall k \in \mathcal{K}\}$ 的线性约束。因此,约束(3-9)以及(3-10)可以被写为如下形式:

$$\sum_{k=1}^K t_k \leq T^{\max} - T_0 \sum_{k=1}^K c_k, \quad (3-13)$$

$$\sum_{k=1}^K t_k \leq \frac{E^{\max}}{P^{\text{com}}} - \frac{P^{\text{sen}}}{P^{\text{com}}} T_0 \sum_{k=1}^K c_k. \quad (3-14)$$

注意到变换后的约束条件(3-13)和(3-14)都是关于子问题中所有用户的总传输时间的线性约束。为了便于表述,本文引入辅助变量 D 和 Q_k :

$$D = \min \left\{ T^{\max} - T_0 \sum_{k=1}^K c_k, \frac{E^{\max}}{P^{\text{com}}} - \frac{P^{\text{sen}}}{P^{\text{com}}} T_0 \sum_{k=1}^K c_k \right\}, \quad (3-15)$$

$$Q_k = \frac{1}{T} B \log_2 \left(1 + \frac{g_k P^{\text{com}}}{\delta} \right). \quad (3-16)$$

文中将问题(3-8)转化为子问题:

$$\text{maximize } \sum_{k=1}^K \ln(\Theta_k(c_k)Q_k t_k), \quad (3-17)$$

$$\text{subject to: } 0 \leq \sum_{k=1}^K t_k \leq D, \quad (3-18)$$

$$\text{variables: } t_k, \forall k \in \mathcal{K}. \quad (3-19)$$

对子问题(3-17)求解, 可以获得 $\{t_k, \forall k \in \mathcal{K}\}$ 。因此, 文中假设已知 $\{t_k, \forall k \in \mathcal{K}\}$, 构建顶层问题, 求解 $\{c_k, \forall k \in \mathcal{K}\}$:

$$\text{maximize } \sum_{k=1}^K \ln(\Theta_k(c_k)Q_k t_k), \quad (3-20)$$

$$\text{subject to: } 0 \leq \sum_{k=1}^K t_k \leq D, \quad (3-21)$$

$$\text{variables: } c_k, \forall k \in \mathcal{K}. \quad (3-22)$$

对目标函数公式参考此处(3-17)求二阶导可证明其总是小于零(证明见附录), 因此可得问题(3-17)是一个对 $\{t_k, \forall k \in \mathcal{K}\}$ 严格凸的问题。因此, 文中采用 KKT 条件 (Karush-Kuhn-Tucker Conditions) [31] 进行求解。首先, 引入辅助变量 λ_k , 构建问题(3-17)的拉格朗日函数[32]:

$$\mathcal{L}(t_k, \lambda) = \sum_{k=1}^K \ln(\Theta_k(c_k)Q_k t_k) + \lambda(D - \sum_{k=1}^K t_k). \quad (3-23)$$

根据拉格朗日函数求一阶偏导为零的条件, 进行推导:

$$\frac{\partial \mathcal{L}(t_k, \lambda)}{\partial t_k} = \frac{1}{\Theta_k(c_k)Q_k} - \lambda = 0, \quad (3-24)$$

$$\frac{\partial \mathcal{L}(t_k, \lambda)}{\partial \lambda} = D - \sum_{k=1}^K t_k = 0, \quad (3-25)$$

可以求得:

$$t_k = \frac{1}{\lambda}, \quad (3-26)$$

$$\sum_{k=1}^K t_k = D. \quad (3-27)$$

综上所述, 子问题(3-17)求解完毕。下一步, 文中写出了在子问题已求解的情况下, 顶层问题带入子问题求解结果后, 相当于原问题(3-8)的等价形式:

$$\text{maximize } \sum_{k=1}^K \ln(\Theta_k(c_k)) + K \ln D, \quad (3-28)$$

$$\text{subject to: } 0 \leq \sum_{k=1}^K t_k \leq D, \quad (3-29)$$

$$\text{variables: } c_k, \forall k \in \mathcal{K}. \quad (3-30)$$

通过，以上转换，可以得知 D 的值取决于：

$$y = \sum_{k=1}^K c_k. \quad (3-31)$$

因此，原文提出一种线性搜索与短视分配策略相结合的顶层问题求解思路。针对 y 进行步进式搜索，每次增加一个感知轮数，根据各个行人的感知准确率曲线，把增加的感知轮数分配给准确率提高最多的行人，计算目标函数的值；记录整个搜索空间内的所有目标函数值以及对应的感知分配策略，取使得目标函数最大的分配策略。

通过对目标函数(3-28)的分析，可得知，在最小值函数确定取第一项或第二项的情况下， D 关于 y 是递减的：

$$D(y) = \min \left\{ T^{\max} - T_0 y, \frac{E^{\max}}{P^{\text{com}}} - \frac{P^{\text{sen}}}{P^{\text{com}}} T_0 y \right\}. \quad (3-32)$$

并且 $\Theta_k(c_k) = \gamma_k - \alpha_k c_k^{-\beta_k}$ 是关于 c_k 的单调递增函数。也就是说，对单个行人来说，分配到的感知轮数越多，准确率越高，然而每个行人的准确率都有各自的渐进上限。换言之，随着 y 的增加， $\sum_{k=1}^K \ln(\Theta_k(c_k))$ 是非递减的。下图所示为 5 个行人随机生成的准确率模型参数随分配的感知轮数变化曲线。

综上，目标函数(3-28)不是关于 y 单调的。为了最大化目标函数，作者提出一种从最小值到最大值的线性搜索算法，通过遍历取值，计算目标函数的值，最终通过取目标函数最大值，来得到最优的 y 。同时，为了最大化 $\sum_{k=1}^K \ln(\Theta_k(c_k))$ ，每次将增加的轮数分配给准确率增加最大的行人。

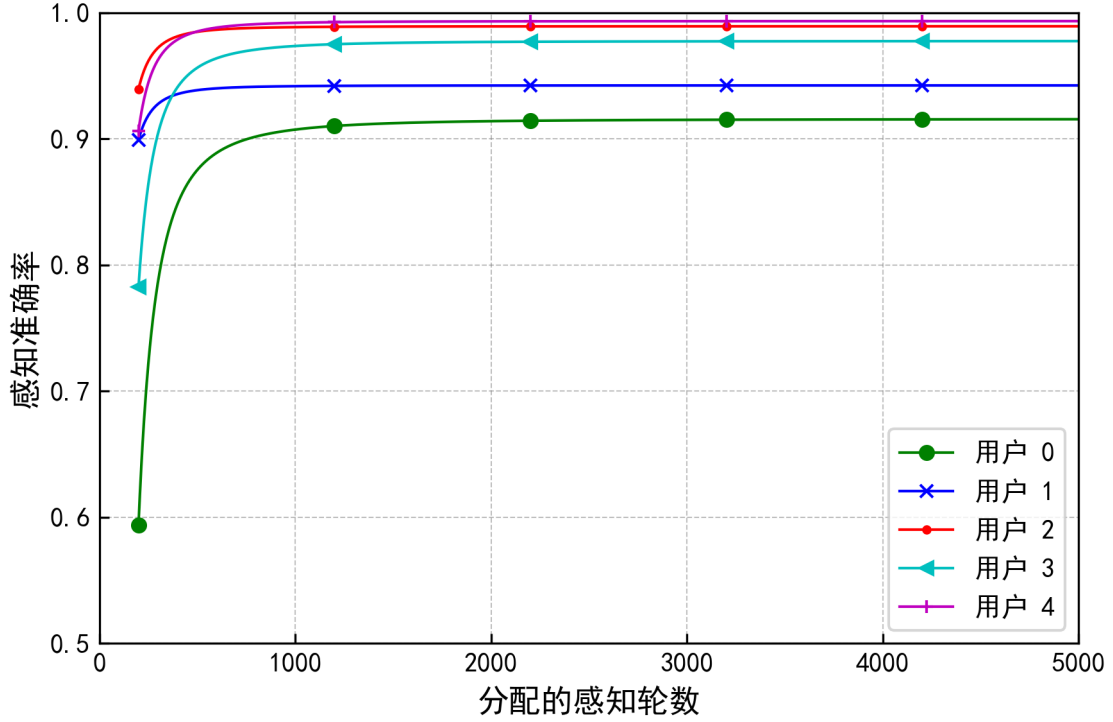


图 3-2 各用户感知准确率与感知轮数的关系曲线

3.4 实验复现结果及分析

本文针对上一章节阐述的时分复用的通感一体化系统做了算法复现以及性能评价。本文的实验平台是一台具有 Intel(R) Core(TM) i5-12500H 的 CPU 的 PC 上进行；主机装有 Windows 11 专业版，实验在适用于 Windows 的 Linux 子系统进行，所使用的系统版本是 Ubuntu 20.04 LTS；复现论文使用的语言是 Python 3.7.10。此外，为复现文中的求解效果对比，将采用 LINGO 20.0 进行求解。

首先，本文根据原论文^[11]中的系统建模以及算法设计进行复现；接着，本文对复现后的算法进行评测，与文中所提出的固定策略算法进行对比。下图所示的固定策略算法具体为：把总时间按比例 ρ , $0 \leq \rho \leq 1$ 分配给感知用户， $(1 - \rho)$ 分配给通信用户，用户之间采用公平分配的策略。

在 5、10、15、20、25、30 个用户的配置下，原文^[11]算法所得到的有效吞吐量和如下所示：

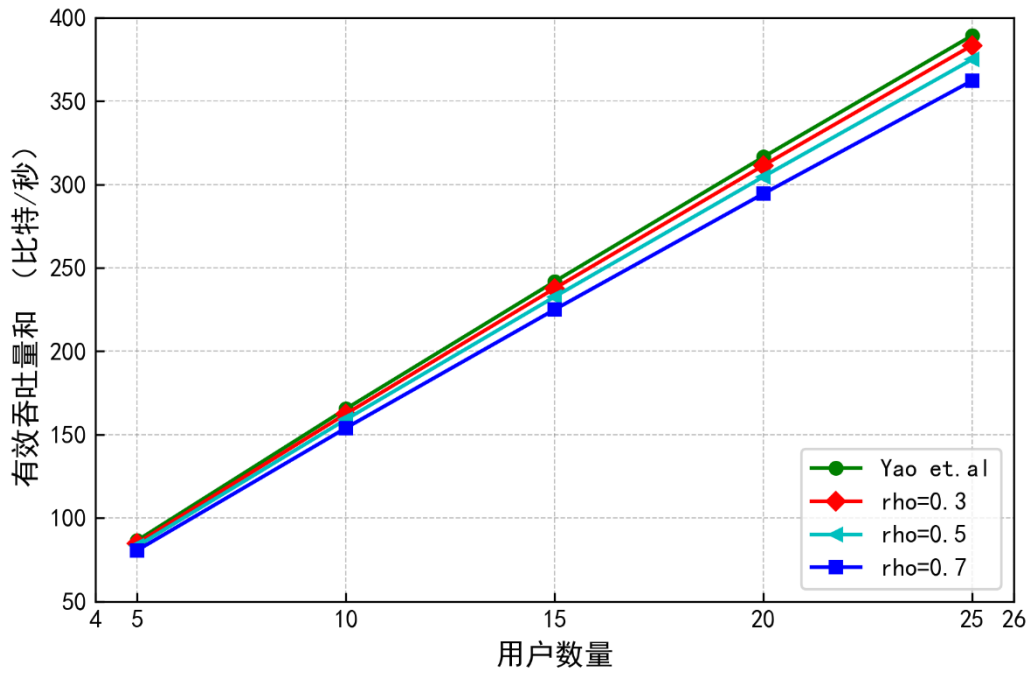


图 3-3 原文算法与基线算法有效吞吐量和的对比

根据上图，可以看出，原文^[11]的效果已超越三种参数下的平均分配策略；同时，在配置范围内，随着用户数量的增加，有效吞吐量和呈上升趋势。

利用求解器 LINGO^[33]对问题(3-8)在 5、10、15、20、25、30 个用户的配置下，采用相同的随机参数配置，进行求解。在本实验中，获得与原论文中相近的表现，作者所提出的算法的有效吞吐量和与 LINGO 解出的全局最优解的性能非常接近。

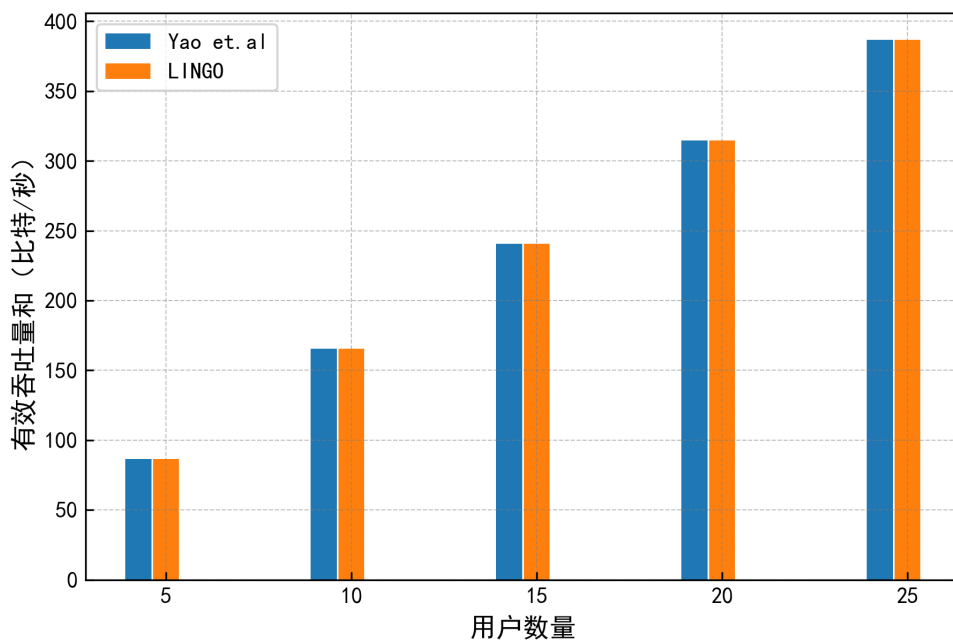


图 3-4 原文算法与 LINGO 求解效果对比

3.5 存在的问题

在假设 ISAC 周期内（指阶段一与阶段二的总时长内），信道不发生变化，且对各个行人的感知准确率模型与实际相符的情况，根据上一节的实验复现结果，可以得知原文^[11]在算法求解性能方面已经非常接近最优。然而，本文存在以下几个亟待改进的问题：

- 假设通信与感知的用户是同一个实体，并且目标函数上，通信与感知是相互耦合的，即先感知再通信。这个假设使得原文^[11]中的算法有失一般性，不具有推广价值。
- 假设机动车部署的通感一体化设备只具有单发天线、单收天线，构建了一个单输入单输出（Single Input Single Output, SISO）系统。然而，随着下一代无线通信技术的不断发展落地，大规模多输入多输出技术（Massive Multiple Input Multiple Output, Massive MIMO）已得到推广^[1]。然而，原文中的场景，无法拓展到 Massive MIMO 的场景下。
- 原文^[11]的感知准确率模型假设有待考量，作者使用感知准确率模型来表征感知的性能，感知准确率仅仅和用户的准确率模型参数配置以及分配到的时域资源量有关，而与频域、带宽、功率等没有建立联系。

针对以上三点不足，在下一章节，本文将设计一个基于空分复用的、多用户多输入多输出（Multi-User Multiple Input Multiple Output System, MU-MIMO System）的通感一体化网络资源调度系统，并采用雷达估计信息率作为感知业务的性能指标。

3.6 本章小结

本章节对一篇以时分复用方式构建的通感一体化资源调度系统的论文^[11]进行算法复现以及效果评测。本章对该资源调度系统的通信吞吐量、感知准确率、有效吞吐量进行了分析评估，并且详细阐述分析结果，探讨其影响因素。此外，本章还将对该资源调度系统的优缺点进行了探讨，包括其在实际应用中可能面临的挑战和限制。本章提出了一系列针对该系统的改进方向和建议，以期进一步提高其性能和实用性。在下一章节中，本文将详细介绍针对上述问题所设计的新的通感一体化资源调度系统的实现细节和优势特点。

4 多用户空分复用的通感一体化网络系统建模

本章节将设计一个多用户、多输入多输出的通感一体化网络资源调度系统，将雷达感知与下行通信相结合，能够在不同的频段上进行雷达感知和下行通信操作，并使用空分复用的方式将多个用户的信号同时传输到接收端，从而提高了通信系统的效率和性能。本章节的主要研究内容包括：对场景、信号、信道进行建模，构建数字信号处理流程，以及评价指标的建模。具体而言，本文将采用雷达估计信息率作为感知业务的性能指标，采用可达速率作为下行通信的性能指标。通过对场景、信号、信道的建模，本文可以更好地了解多用户空分复用的通感一体化网络系统的工作原理和性能，进而在下一章节构建优化系统的调度策略。

4.1 场景介绍

首先，本文假设系统中存在一共 K 个用户，其中共 L 个雷达感知用户，共 M 个下行通信用户。本章用 $\mathcal{K} = \{1, \dots, K\}$ 表示全体用户集合，用 $\mathcal{M} = \{1, \dots, M\}$ 表示通信用户集合，用 $\mathcal{L} = \{1, \dots, L\}$ 表示感知用户的集合。一个 DFRC 基站 (Dual-functional Radar and Communication Base Station, DFRC BS) 通过发射通信信号波形向通信用户提供下行通信服务，通过发射雷达探测波形继而接收其雷达回波，对感知用户进行测量（如测量距离、方位、速度等）。

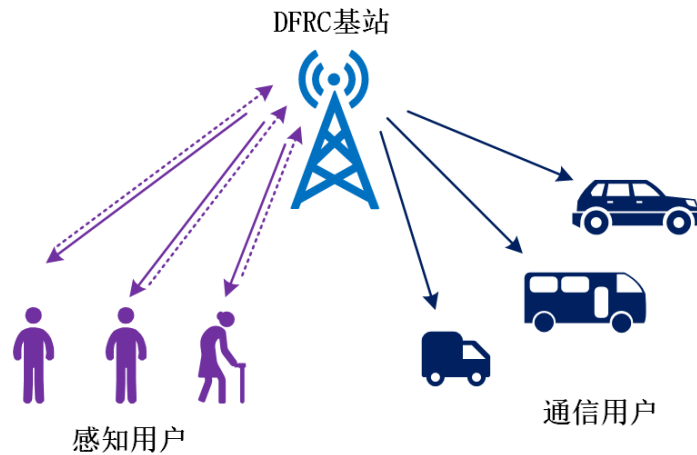


图 4-1 多用户空分复用的通感一体化系统场景示意图

为了避免通信业务以及感知业务之间的干扰，下行通信业务和感知业务分别处于两个不交叠的频段，频带宽度分别为 B_c 以及 B_r 。

DFRC 基站配备有 N_t 个发天线单元以及同等数量个收天线单元，用于发射波形以及接收雷达回波信号。每个用户配备 ($N_r = 1$) 个接收天线单元，用于接收下行通信信号。本文假设 $N_t \leq MN_r$ ，以确保用户之间的信号干扰可以被消除。

4.2 下行通信信号模型

对于通信用户 $m \in \mathcal{M}$ ，其接收到的下行通信信号可表达为：

$$\mathbf{y}_m = \sqrt{p_m^c} \mathbf{h}_m \mathbf{w}_m^c s_m^c + \mathbf{h}_m \sum_{k \in \mathcal{M}, k \neq m} \sqrt{p_k^c} \mathbf{w}_k^c s_k^c + \mathbf{n}_c. \quad (4-1)$$

其中， p_m^c 为基站对通信用户 m 的下行通信发射功率， $\mathbf{h}_m \in \mathbb{C}^{N_r \times N_t}$ 为通信用户 m 的下行信道矩阵， s_m^c 为发送的信号， $\mathbf{w}_m^c \in \mathbb{C}^{N_t \times 1}$ 为用户 m 的预编码矩阵， \mathbf{n}_c 为加性高斯白噪声。对于通信用户 m ，第一项为有用信号，第二、三项为干扰项以及噪声项。下行信道的具体建模为：

$$\mathbf{h}_m^c = \sqrt{g_m} \sqrt{N_t N_r} \mathbf{a}_r(\theta_l) \mathbf{a}_t(\theta_l)^H. \quad (4-2)$$

上式中， g_m 为路径损耗， $\mathbf{a}_r(\theta_l)$ 与 $\mathbf{a}_t(\theta_l)$ 分别为接收端以及发送端在方向角 θ_l 上的导向矢量，其表达形式如下：

$$\mathbf{a}_t(\theta_l) = \frac{1}{\sqrt{N_t}} [1, \dots, e^{j\pi(N_t-1)\sin\theta_l}]^T, \quad (4-3)$$

$$\mathbf{a}_r(\theta_l) = \frac{1}{\sqrt{N_r}} [1, \dots, e^{j\pi(N_r-1)\sin\theta_l}]^T. \quad (4-4)$$

根据 Friis 传输公式^[34]，路径损耗可表示为距离与信号频率的函数：

$$g_m = \left(\frac{c}{4\pi d_m f} \right)^2. \quad (4-5)$$

4.3 感知信号模型

根据^[10]，多普勒频移可以通过同步手段消除其影响，进而进行后续的数字信号处理流程。因为感知业务工作在另一个频带，不受通信信号的干扰，基站收到的关于感知用户 l 的雷达回波信号可以表示为：

$$\begin{aligned} \mathbf{y}_l = & \sqrt{g_l^r} N_t \mathbf{a}_t(\theta_l) \mathbf{a}_t(\theta_l)^H \mathbf{x}(t - \tau_l) \\ & + \sum_{p \in \mathcal{L}, p \neq l} \sqrt{g_p^r} N_t \mathbf{a}_t(\theta_p) \mathbf{a}_t(\theta_p)^H \mathbf{x}(t - \tau_p) + \sum_{k \in \mathcal{M}} \sqrt{g_k^r} N_t \mathbf{a}_t(\theta_k) \mathbf{a}_t(\theta_k)^H \mathbf{x}(t - \tau_k) + \mathbf{n}_r. \end{aligned} \quad (4-6)$$

其中, 第一项为有用回波信号, 第二项为其他感知用户反射的回波信号, 第三项为通信用户反射的回波干扰信号, 第四项 \mathbf{n}_l 为加性高斯白噪声信号。 g_l^r 是雷达回波的路径衰落。 \mathbf{x} 为基站发射的雷达信号为各感知用户对应发射信号的线性和。其中 $\mathbf{v}_l \in \mathbb{C}^{N_t \times 1}$ 是感知用户 l 的预编码矩阵, p_l^r 是感知用户 l 雷达信号的发射功率。

$$\mathbf{x} = \sum_{l \in \mathcal{L}} \mathbf{v}_l \sqrt{p_l^r} s_l. \quad (4-7)$$

\mathbf{v}_l 可以采用在估计的方向角上的 $\hat{\theta}_l$ 发送导向矢量^[35]:

$$\mathbf{v}_l = \mathbf{a}_t(\hat{\theta}_l). \quad (4-8)$$

根据^[34], 雷达回波信道路径衰落表达式如下:

$$g_l^r = \left(\frac{c}{4\pi d_l^2 f} \right)^2. \quad (4-9)$$

4.4 多用户通信信号检测模型

根据(4-1)可知, 对于通信用户 $m \in \mathcal{M}$, 接收到的下行通信信号表达式:

$$\mathbf{y}_m = \underbrace{\sqrt{p_m^c} \mathbf{h}_m \mathbf{w}_m^c s_m^c}_{\text{useful signal}} + \underbrace{\mathbf{h}_m \sum_{k \in \mathcal{M}, k \neq m} \sqrt{p_k^c} \mathbf{w}_k^c s_k^c}_{\text{inter-user interference}} + \mathbf{n}_c. \quad (4-10)$$

定义扩展的信道矩阵 $\mathbf{H} \in \mathbb{C}^{M \times N_t}$ 为:

$$\mathbf{H} \triangleq [\mathbf{h}_1, \dots, \mathbf{h}_M]^T. \quad (4-11)$$

定义扩展的预编码矩阵 $\mathbf{W}_c \in \mathbb{C}^{N_t \times M}$ 为:

$$\mathbf{W}_c \triangleq [\mathbf{w}_1, \dots, \mathbf{w}_M]. \quad (4-12)$$

定义扩展的信号矩阵 $\mathbf{s} \in \mathbb{C}^{M \times 1}$ 为 $\mathbf{s} = [p_1^c s_1^c, \dots, p_M^c s_M^c]^T$ 。下行接收信号 $\mathbf{y} \in \mathbb{C}^{M \times 1}$ 可写为:

$$\mathbf{y} = \mathbf{H} \mathbf{W}_c \mathbf{s} + \mathbf{n}_c. \quad (4-13)$$

迫零预编码 (Zero Forcing, ZF) 是一种常用的下行多用户预编码技术。迫零预编码通过预编码矩阵, 迫使其他用户的干扰为零。本文假设 $N_t \leq MN_r$, 以使得用户之间的信号干扰可以被消除。根据^[36], 可知扩展的预编码矩阵表达式为:

$$\mathbf{W}_c = \beta \mathbf{H}^H (\mathbf{H} \mathbf{H}^H)^{-1}, \quad (4-14)$$

$$\beta = \sqrt{\frac{1}{\|\mathbf{H}\mathbf{H}^H\|_F^2}}. \quad (4-15)$$

将(4-14)代入(4-13)，可得出下行接收信号中的多通信用户之间的干扰信号被消除。

$$\begin{aligned} \mathbf{y} &= \beta \mathbf{H}\mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1} \mathbf{s} + \mathbf{n}_c \\ &= \beta \mathbf{s} + \mathbf{n}_c. \end{aligned} \quad (4-16)$$

4.5 通信性能指标

根据(4-10)，通信用户 m 的信干噪比为：

$$\gamma_m^c = \frac{p_m^c |\mathbf{h}_l \mathbf{w}_l|^2}{\sum_{q \in \mathcal{M}, q \neq m} p_q^c |\mathbf{h}_q \mathbf{w}_q|^2 + \sigma_c^2}. \quad (4-17)$$

也可以根据(4-16)写为：

$$\gamma_m^c = \frac{\beta p_m^c}{\sigma_c^2}. \quad (4-18)$$

通信用户 m 的下行可达速率为：

$$R_m^c = B_c \log_2(1 + \gamma_m^c). \quad (4-19)$$

所有通信用户的下行可达和速率为：

$$R_{sum}^c = \sum_{m \in \mathcal{M}} B_c \log_2(1 + \gamma_m^c). \quad (4-20)$$

4.6 多目标雷达信号检测模型

因为感知业务工作在另一个频带，不受通信信号的干扰，基站收到的关于感知用户 l 的雷达回波信号含有有用回波信号、杂波干扰信号（Clutter）及噪声：

$$\begin{aligned} \mathbf{y}_l(t) &= \underbrace{\sqrt{g_l^r} N_l \mathbf{a}_l(\theta_l) \mathbf{a}_l(\theta_l)^H \mathbf{x}(t - \tau_l)}_{\text{useful signal}} \\ &+ \underbrace{\sum_{p \in \mathcal{L}, p \neq l} \sqrt{g_p^r} N_p \mathbf{a}_p(\theta_p) \mathbf{a}_p(\theta_p)^H \mathbf{x}(t - \tau_p) + \sum_{k \in \mathcal{M}} \sqrt{g_k^r} N_k \mathbf{a}_k(\theta_k) \mathbf{a}_k(\theta_k)^H \mathbf{x}(t - \tau_k)}_{\text{clutter signal}} + \mathbf{n}_r. \end{aligned} \quad (4-21)$$

此处假设同一个用户的雷达波形的时域自相关为 1，不同的用户发射的波形不相关求得 0，如下：

$$\begin{aligned} \int s_l(t)s_l^*(t-\tau_l) dt &= 1, \forall l \in \mathcal{L}, \\ \int s_l(t)s_k^*(t-\tau_k) dt &= 0, \forall l, k \in \mathcal{L}, k \neq l. \end{aligned} \quad (4-22)$$

对接收到的信号使用匹配滤波器^[37], 可得感知用户 l 匹配后的信号表达式为:

$$\tilde{y}_l = \int \mathbf{y}(t)s_l^*(t-\tau_l) dt, \quad (4-23)$$

$$\begin{aligned} \tilde{\mathbf{y}}_l &= \sqrt{p_l^r} \sqrt{g_l^r} N_l \mathbf{a}_l(\theta_l) \mathbf{a}_l(\theta_l)^H \mathbf{v}_l + \sum_{p \in \mathcal{L}, p \neq l} \sqrt{p_l^r} \sqrt{g_p^r} N_l \mathbf{a}_l(\theta_p) \mathbf{a}_l(\theta_p)^H \mathbf{v}_l \\ &\quad + \sum_{k \in \mathcal{M}} \sqrt{p_l^r} \sqrt{g_k^r} N_l \mathbf{a}_l(\theta_k) \mathbf{a}_l(\theta_k)^H \mathbf{v}_l + \mathbf{n}_r. \end{aligned} \quad (4-24)$$

最小方差无失真响应波束成形器 (Minimum Variance Distortionless Response Beamforming, MVDR Beamforming) 是一种基于最小均方误差准则的滤波器^[38]。在约束目标方向的增益无畸变的情况下, 使得干扰方向上的信号最小, 从而达到干扰抑制的效果。MVDR 波束成形器广泛应用于毫米波雷达的方向角估计、杂波干扰抑制等方面^[38]。本文使用 MVDR 滤波器, 抑制杂波干扰。定义感知用户 l 的接收矩阵 \mathbf{u}_l , 则匹配滤波后的信号可以进一步处理为:

$$\begin{aligned} \mathbf{r}_l &= \mathbf{u}_l^H \tilde{\mathbf{y}}_l \\ &= \sqrt{p_l^r} \sqrt{g_l^r} N_l \mathbf{u}_l^H \mathbf{a}_l(\theta_l) \mathbf{a}_l(\theta_l)^H \mathbf{v}_l + \sum_{p \in \mathcal{L}, p \neq l} \sqrt{p_l^r} \sqrt{g_p^r} N_l \mathbf{u}_l^H \mathbf{a}_l(\theta_p) \mathbf{a}_l(\theta_p)^H \mathbf{v}_l \\ &\quad + \sum_{k \in \mathcal{M}} \sqrt{p_l^r} \sqrt{g_k^r} N_l \mathbf{u}_l^H \mathbf{a}_l(\theta_k) \mathbf{a}_l(\theta_k)^H \mathbf{v}_l + \mathbf{u}_l^H \mathbf{n}_r. \end{aligned} \quad (4-25)$$

本文定义 $\mathbf{A}(\theta_l) = \mathbf{a}_l(\theta_l) \mathbf{a}_l(\theta_l)^H$, 则信号 \mathbf{r}_l 的干扰加噪声项的相关矩阵写为:

$$\begin{aligned} \mathbf{R}_l &= \sum_{q \in \mathcal{L}, q \neq l} p_l^r g_q^r N_l^2 \mathbf{A}(\theta_q) \mathbf{v}_l \mathbf{v}_l^H \mathbf{A}(\theta_q)^H \\ &\quad + \sum_{k \in \mathcal{M}} p_l^r g_k^r N_l^2 \mathbf{A}(\theta_k) \mathbf{v}_l \mathbf{v}_l^H \mathbf{A}(\theta_k)^H + \sigma_r^2 \mathbf{I}. \end{aligned} \quad (4-26)$$

根据^[39], 感知用户 l 的接收矩阵 \mathbf{u}_l 表达式为:

$$\mathbf{u}_l = \frac{\mathbf{R}_l^{-1} \mathbf{A}(\theta_l) \mathbf{v}_l}{\mathbf{v}_l^H \mathbf{A}(\theta_l)^H \mathbf{R}_l^{-1} \mathbf{A}(\theta_l) \mathbf{v}_l}. \quad (4-27)$$

4.7 感知性能指标

根据(4-25), 感知用户 l 的信干噪比可以表示为:

$$\gamma_l^r = \frac{p_l^r g_l^r N_l^2 \|\mathbf{u}_l^H \mathbf{A}(\theta_l) \mathbf{v}_l\|_F^2}{\sum_{p \in \mathcal{L}, p \neq l} p_l^r g_p^r N_l^2 \|\mathbf{u}_l^H \mathbf{A}(\theta_p) \mathbf{v}_l\|_F^2 + \sum_{k \in \mathcal{M}} p_l^r g_k^r N_l^2 \|\mathbf{u}_l^H \mathbf{A}(\theta_k) \mathbf{v}_l\|_F^2 + \sigma_r^2 \|\mathbf{u}_l^H\|_F^2}. \quad (4-28)$$

也可以写为:

$$\gamma_l^r = p_l^r \mathbf{v}_l^H (\mathbf{h}_l^r)^H \mathbf{R}_l^{-1} \mathbf{h}_l^r \mathbf{v}_l, \quad (4-29)$$

$$\mathbf{h}_l^r = \sqrt{g_l^r} N_l \mathbf{a}_l(\theta_l) \mathbf{a}_l(\theta_l)^H. \quad (4-30)$$

本文将采用雷达信息估计率 (Radar Estimation Information Rate) ^[40] 作为感知性能的衡量标准。其中, δ 是雷达波形占空比, σ_{proc}^2 是感知目标的距离误差。

$$R_l^r = \frac{\delta}{2T} B_r \log_2 \left(1 + 4\pi^2 \sigma_{proc}^2 B_r^2 T \gamma_l^r \right). \quad (4-31)$$

总雷达信息估计速率为:

$$R_{sum}^r = \sum_{l \in \mathcal{L}} R_l^r. \quad (4-32)$$

雷达信息估计率指标类似于通信系统中的数据信息速率。雷达发射端对目标的照射雷达波形的过程, 可以看作是目标被动地向雷达传递其参数信息, 那么雷达信道可以被描述为“不合作”的通信信道。雷达信息估计率, 则可以被看作是雷达和目标之间的互信息 (Mutual Information), 综合考虑被估计的随机参数的熵和该参数的估计不确定性的熵, 来共同构建雷达信息估计率^[40]。

4.8 本章小结

本章设计了一个基于空分复用的、多用户多输入多输出 (Multi-User Multiple Input Multiple Output, MU-MIMO) 的通感一体化网络资源调度系统。雷达感知与下行通信工作在不同的频段; 而在同一业务中, 用户之间以空分复用的方式工作在同一频带。本文采用雷达估计信息率作为感知业务的性能指标, 采用可达速率作为下行通信的性能指标。本章节完成了对于场景、信号、信道的建模、数字信号处理流程的构建, 以及评价指标的建模。

5 基于 PPO 的多用户空分复用场景下的 资源调度算法设计与实现

本章将基于上一章节所提出的场景，构建一个基于强化学习的多用户空分复用场景下的功率分配问题。接着，本文将基于问题建模，提出合理的资源调度系统设计方案，并阐述算法的实现流程。

5.1 问题建模

本文构建了一个多用户通感一体化系统，这个系统中存在一共 K 个用户，其中共 L 个雷达感知用户，共 M 个下行通信用户。一个 DFRC 基站通过发射通信信号波形向通信用户提供下行通信服务，通过发射雷达探测波形继而接收其雷达回波，对感知用户进行测量。下行通信业务和感知业务分别处于两个不交叠的频段，频带宽度分别为 B_c 以及 B_r ，因此两个业务之间不会产生干扰。DFRC 基站配备有 N_t 个发天线单元以及同等数量个收天线单元，用于发射波形以及接收雷达回波信号。每个用户配备 ($N_r = 1$) 个接收天线单元，用于接收下行通信信号。

在上一章节，本文对下行通信过程、雷达感知过程做了详细的数学建模；本文采用迫零预编码，通过发掘空域的自由度，从而抑制多通信用户之间干扰，使得多个通信用户能够共享频带；在一个多目标雷达感知系统中，多个目标的检测波形会共享整个频带；然而，不仅是目标感知用户会反射其对应的波形，其他感知用户也会反射该用户的波形，与此同时，通信用户也会反射感知波形，造成杂波干扰；因此，本文采用 MVDR 波束成形器来抑制干扰方向的来波，从而实现多感知用户的同频雷达感知。

本文以加权的雷达估计信息率与通信用户的下行可达和速率作为目标函数，该函数是通信功率向量 $\mathbf{p}_c = [p_1^c, \dots, p_M^c]$ 以及感知功率向量 $\mathbf{p}^r = [p_1^r, \dots, p_L^r]$ 的函数。

$$\begin{aligned} \Theta(\mathbf{p}^c, \mathbf{p}^r) &= \omega_1 R_{sum}^c(\mathbf{p}^c) + \omega_2 R_{sum}^r(\mathbf{p}^r) \\ &= \omega_1 \sum_{m \in \mathcal{M}} B_c \log_2(1 + \gamma_m^c) + \omega_2 \sum_{l \in \mathcal{L}} \frac{\delta}{2T} B_r \log_2(1 + 4\pi^2 \sigma_{proc}^2 B_r^2 T \gamma_l^r). \end{aligned} \quad (5-1)$$

对于该通感一体化系统中，总功率有一个上限约束，因此，各个通信用户、感知用户的总功率和应满足小于等于该上限。

$$\sum_{m \in \mathcal{M}} p_m^c + \sum_{l \in \mathcal{L}} p_l^r \leq P^{\text{tot}}. \quad (5-2)$$

对于单个通信或者感知用户分配到的功率而言，应有上限与下限的约束。

$$0 \leq p_m^c \leq P^{\max}, \forall m \in \mathcal{M}. \quad (5-3)$$

$$0 \leq p_l^r \leq P^{\max}, \forall l \in \mathcal{L}. \quad (5-4)$$

可以推导出如下问题建模：

$$\text{maximize} \quad \Theta(\mathbf{p}^c, \mathbf{p}^r) = \omega_1 R_{\text{sum}}^c(\mathbf{p}^c) + \omega_2 R_{\text{sum}}^r(\mathbf{p}^r), \quad (5-5)$$

$$\sum_{m \in \mathcal{M}} p_m^c + \sum_{l \in \mathcal{L}} p_l^r \leq P^{\text{tot}}, \quad (5-6)$$

$$0 \leq p_m^c \leq P^{\max}, \forall m \in \mathcal{M}, \quad (5-7)$$

$$0 \leq p_l^r \leq P^{\max}, \forall l \in \mathcal{L}, \quad (5-8)$$

$$\text{variables:} \quad \mathbf{p}^c, \mathbf{p}^r. \quad (5-9)$$

由于迫零预编码后的通信用户的信道被均衡，在给定通信用户功率和的情况下，最大的通信用户速率和应为功率在通信用户之间平均分配的情况：

$$\begin{aligned} \sum_{m \in \mathcal{M}} p_m^c &\leq P_c^{\text{tot}}, \\ \max R_m^c(\mathbf{p}^c) &= MB_c \log_2 \left(1 + \frac{\beta P_c^{\text{tot}}}{M \sigma_c^2} \right). \end{aligned} \quad (5-10)$$

综上，问题(5-5)可以被转换为如下等价形式：

$$\text{maximize} \quad \Theta(\alpha, \boldsymbol{\beta}) = \omega_1 R_{\text{sum}}^c(\alpha) + \omega_2 R_{\text{sum}}^r(\boldsymbol{\beta}), \quad (5-11)$$

$$\alpha P^{\text{tot}} + \sum_{l \in \mathcal{L}} \beta_l P^{\text{tot}} \leq P^{\text{tot}}, \quad (5-12)$$

$$0 \leq \beta_l \leq 1, \forall l \in \mathcal{L}, \quad (5-13)$$

$$0 \leq \alpha \leq 1, \quad (5-14)$$

$$\text{variables:} \quad \alpha, \boldsymbol{\beta} = [\beta_1, \dots, \beta_L]. \quad (5-15)$$

本文将问题(5-5)优化每一个用户的通信与感知变量，转换成优化感知用户的能量分配比例，余下的能量则被通信用户平分。经过转换之后，原功率分配问题转化为分配比例优化问题，但是转化后的目标函数关于分配比例 $\boldsymbol{\beta} = [\beta_1, \dots, \beta_L]$ 仍然是一个非凸的问题。因此，本文设计了一种基于强化学习算法近端策略优化(Proximal Policy Optimization)的功率分配算法，来求解问题(5-5)。

5.2 马尔可夫决策过程设计

近端策略优化算法 (Proximal Policy Optimization Algorithm, PPO Algorithm)^[41]是信赖域策略优化 (Trust Region Policy Optimization Algorithm, TRPO Algorithm)^[42]的改进版本。这两个算法都关注同样的问题：利用目前拥有的数据记录，算法每次采取使得策略进步最大的动作，并防止策略更新过大导致的性能坍缩。TRPO 系列算法是采用复杂的二阶方法，来应对此类问题；而 PPO 系列算法采用一阶方法，技来保持新策略与旧策略的接近。因此，相比于 TRPO 系列算法，PPO 算法从实现角度来看，是一种更加简单可行且有效的策略^[43]。

采用 PPO 算法解决本文提出的通感一体化场景建模中的功率分配问题，有下列优势：

- PPO 算法是一种可在线训练的强化学习算法，这意味着使用基于 PPO 的深度强化学习策略网络，可以在不断变化的信道环境中随时更新策略网络，以适应信道环境的变化。
- PPO 算法是一种非常典型的智能体——评论家 (Actor-Critic) 的强化学习算法，它对于连续动作有着非常优秀的支持；

首先，本文将问题(5-11)构建为一个马尔可夫决策过程 (Markov Decision Process, MDP)^[28]。在时间 t ，PPO 智能体根据当前状态观测，利用策略网络 π_θ 输出动作 a_t ，并在环境中执行 a_t ，获得环境给予的奖励 r_t 。

- 状态设计：在调度时间 t ，DFRC 基站获得当前的雷达感知信道的估计、下行通信信道的估计，表达式为：

$$\hat{\mathbf{H}}_c(t) = [\hat{\mathbf{h}}_1^c(t), \dots, \hat{\mathbf{h}}_M^c(t)]. \quad (5-16)$$

$$\hat{\mathbf{H}}_r(t) = [\hat{\mathbf{h}}_1^r(t), \dots, \hat{\mathbf{h}}_L^r(t)]. \quad (5-17)$$

其中，雷达感知信道的估计来源于对方向角的估计：

$$\hat{\mathbf{h}}_l^r = \sqrt{g_l^r} N_l \mathbf{a}_l(\hat{\theta}_l) \mathbf{a}_l(\hat{\theta}_l)^H. \quad (5-18)$$

PPO 算法的输入状态表达式为：

$$s_t = [\hat{\mathbf{H}}_c(t), \hat{\mathbf{H}}_r(t)]. \quad (5-19)$$

- 动作设计：在调度时间 t ，PPO 智能体根据观测到的各个用户的状态 s_t ，针对于通信与感知用户，进行功率分配，功率分配因子为动作：

$$a_t = [\alpha, \beta_1, \dots, \beta_L]. \quad (5-20)$$

之后，环境根据决策到的 α 与 β ，把 αP^{tot} 的功率均分给 M 个通信用户，把 $[\beta_1 P^{\text{tot}}, \dots, \beta_L P^{\text{tot}}]$ 对应分配给 L 个感知用户。

- 奖励设计：在调度时间 t ，根据决策到的 α 和 β ，进行功率分配，并计算目标函数(5-11)的值，作为正向奖励；(5-11)中的常数 ω_1 、 ω_2 是用来保证 R_{sum}^r 和 R_{sum}^c 在合理的数量级（绝对值在 50 以内），防止网络因奖励产生的更新过大，产生梯度爆炸现象。同时，约束(5-12)将引入一个惩罚项，在满足这个约束的情况下（也就是总功率不超出限制），智能体可以获得正奖励，否则将获得负惩罚。约束(5-13)(5-14)则采用归一化 PPO 智能体网络的输出满足，具体设计将在下一小节得到阐释。综上，智能体的奖励函数设计如下：

$$r_t = r_{obj} + r_{pen}. \quad (5-21)$$

$$r_{pen} = \begin{cases} -\mu, & \alpha + \sum_{l \in \mathcal{L}} \beta_l > 1, \\ 0, & \alpha + \sum_{l \in \mathcal{L}} \beta_l \leq 1. \end{cases} \quad (5-22)$$

5.3 算法流程

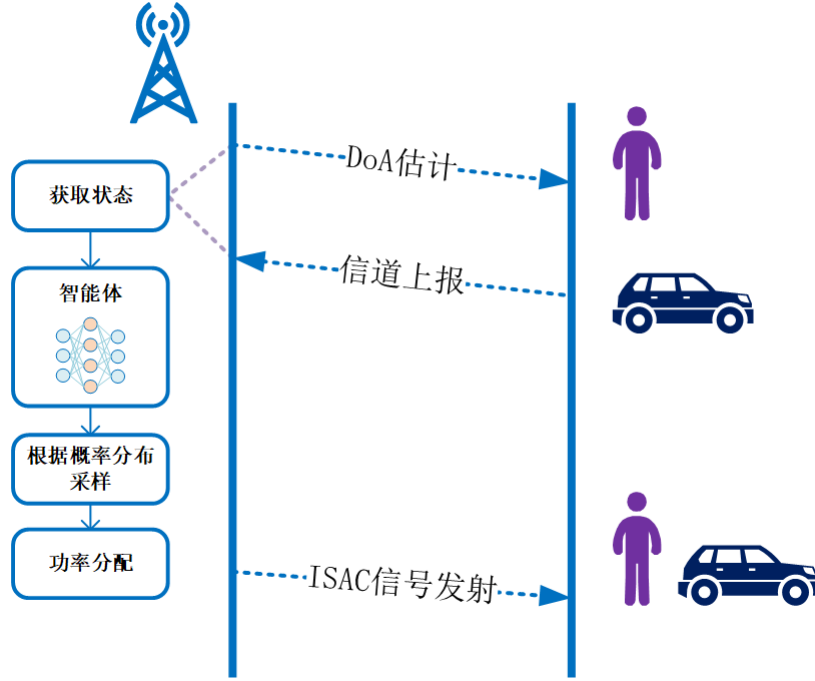


图 5-1 算法流程示意图

首先，本文将阐述一次智能体动作的全部流程，接着本文将阐述基于 PPO 的功率分配网络在本场景中的具体更新流程。

在时间 t ，DFRC 基站通过波达方向估计（Direction of Arrival, DoA）^[37]与接收信号功率测量，获得对感知信道的估计 $\hat{\mathbf{H}}_r(t)$ ，通过通信用户上报测量信号^[44]，获得对下行信道的估计 $\hat{\mathbf{H}}_c(t)$ 。这是智能体获取当前状态的阶段。

接下来，由策略网络输出当前动作，这个动作是由网络输出动作的均值与方差，继而利用多维正态分布函数采样获得对应的功率比例。

$$a_t = \pi_{\theta_t}(s_t). \quad (5-23)$$

接着，本文计算惩罚函数 r_{pen} 。为了使得策略网络输出的动作不违反功率约束 (5-12)，需要将动作 a_t 归一化后，接下来再进行功率分配。

$$\tilde{\alpha}(t) = \frac{\alpha(t)}{\alpha(t) + \sum_{l \in \mathcal{L}} \beta_l(t)}. \quad (5-24)$$

$$\tilde{\beta}_l(t) = \frac{\beta_l(t)}{\alpha(t) + \sum_{l \in \mathcal{L}} \beta_l(t)}. \quad (5-25)$$

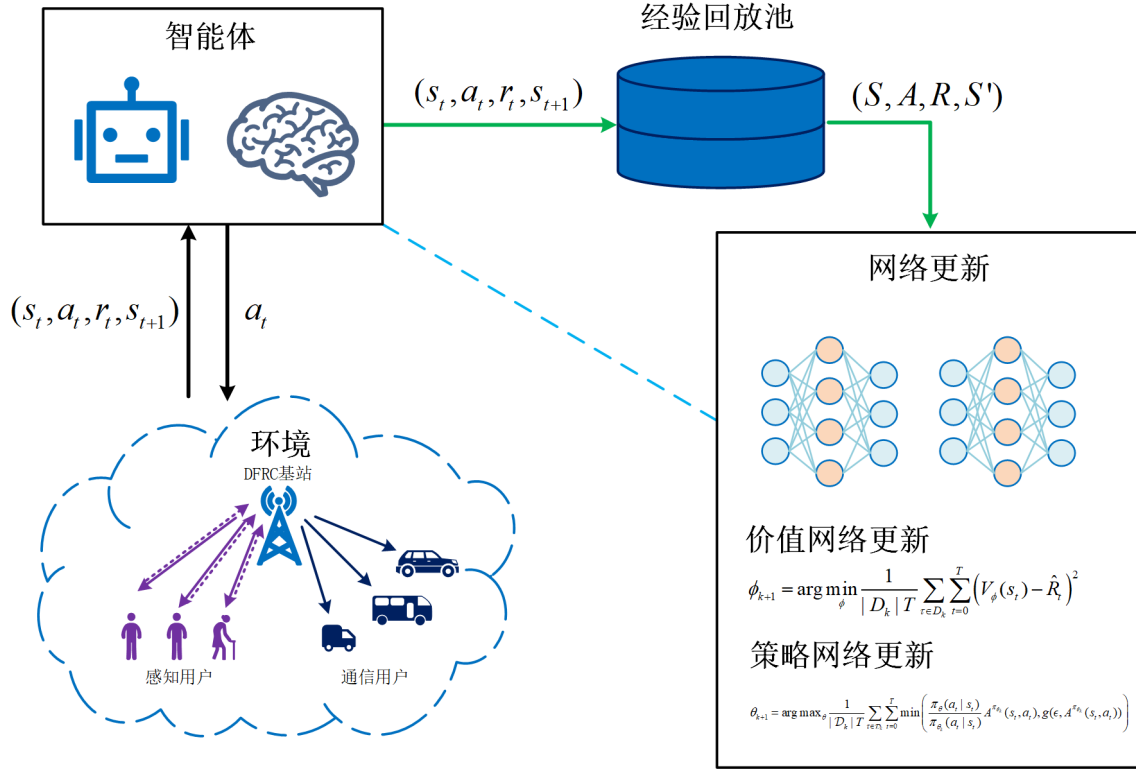


图 5-2 算法训练更新流程示意图

基于 PPO-Clip 算法^[45]的通感一体化网络功率分配算法含有两个神经网络。一个是策略网络 $\pi_{\theta}(s_t)$ ，负责输出功率分配策略；另一个网络是价值网络 $V_{\phi}(s_t)$ ，用来拟合奖励函数。

在每一轮次，智能体策略网络都会根据当前状态 s_t 获取动作 a_t ，并获取下一个状态 s_{t+1} 。并将 (s_t, a_t, r_t, s_{t+1}) 存入经验回放池 \mathcal{D}_k 。

接着，根据价值函数网络的输出估计优势函数，广义优势估计（General Advantage Estimation, GAE）^[46]是一个常用的方法，具体表达如下。

$$\hat{A}_t^{GAE(\gamma, \lambda)} = \sum_{l=0}^{\infty} (\gamma \lambda)^l \delta_{t+l}^V. \quad (5-26)$$

$$\delta_t^V = r_t + \gamma V_{\phi}(s_{t+1}) - V_{\phi}(s_t). \quad (5-27)$$

策略网络的更新基准是：

$$\theta_{k+1} = \arg \max_{\theta} \mathbb{E}_{s, a \sim \pi_{\theta_k}} [L(s, a, \theta_k, \theta)]. \quad (5-28)$$

$$L(s, a, \theta_k, \theta) = \min \left(\frac{\pi_{\theta}(a | s)}{\pi_{\theta_k}(a | s)} A^{\pi_{\theta_k}}(s, a), g(\epsilon, A^{\pi_{\theta_k}}(s, a)) \right). \quad (5-29)$$

$$g(\epsilon, A) = \begin{cases} (1+\epsilon)A, & A \geq 0, \\ (1-\epsilon)A, & A < 0. \end{cases} \quad (5-30)$$

表 5-1 算法更新流程伪代码

算法 1: 基于 PPO-Clip 的通感一体化网络功率分配算法更新流程

Input: 观测环境的初始状态 s_0 ; 初始化策略网络参数 θ_0 , 价值网络参数 ϕ_0

For $k = 0, 1, 2, \dots$ **do**

 在环境中根据当前状态 s_t 运行策略网络动作 $a_t = \pi_{\theta_k}(s_t)$ 获得下一个状态 s_{t+1} 、奖励函数 r_t , 并存于回放池 \mathcal{D}_k

 根据价值函数网络 V_{ϕ_k} , 估计优势值 \hat{A}_t

 采用 Adam 优化器梯度上升更新策略网络:

$$\theta_{k+1} = \arg \max_{\theta} \frac{1}{|\mathcal{D}_k| T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^T \min \left(\frac{\pi_{\theta}(a_t | s_t)}{\pi_{\theta_k}(a_t | s_t)} A^{\pi_{\theta_k}}(s_t, a_t), g(\epsilon, A^{\pi_{\theta_k}}(s_t, a_t)) \right).$$

 采用最小均方误差的损失函数对价值网络进行梯度下降更新:

$$\phi_{k+1} = \arg \min_{\phi} \frac{1}{|D_k| T} \sum_{\tau \in D_k} \sum_{t=0}^T (V_{\phi}(s_t) - \hat{R}_t)^2.$$

End For

5.4 本章小结

在通感一体化网络中, 多用户空分复用技术与多目标检测技术, 使得 DFRC 基站可以同时服务于多个通信用户、同时感知多个感知用户, 从而提高系统的整体性能。在该系统中, 功率分配是一个关键的问题。本章利用强化学习算法, 提出一种合理的资源调度系统设计方案, 并阐述了算法的具体实现流程。

6 仿真与性能评测

本章将对上一章节提出的基于 PPO 的多用户空分复用的资源调度场景下的功率分配算法进行仿真实验,从通信吞吐量、感知吞吐量、加权吞吐量等方面验证其性能。此外,本文还会提出一种基于功率平均分配策略的基线算法进行性能对比。

6.1 仿真场景设置

本章的实验在配备有 Intel(R) Xeon(R) Gold 型号的 CPU 的服务器上进行。该服务器装配 Linux 的系统环境,系统版本是 Ubuntu 20.04 LTS。本章利用该服务器装配的 Nvidia GeForce RTX 3090 显卡加速模型训练,对应使用的 CUDA 版本为 11.7。

本实验在 Python 3.10.4 的语言环境下进行,使用 Miniconda 3 进行 Python 开发环境配置与包管理。本章的非神经网络矩阵运算过程在 CPU 侧执行,采用 NumPy 1.17 的系列库函数实现。本章的神经网络部分,借助于深度神经网络实现框架 PyTorch 1.12.1 实现。

表 6-1 本章实验软硬件配置

软硬件	具体配置
CPU	Intel(R) Xeon(R) Gold
GPU	Nvidia GeForce RTX 3090
系操作统	Ubuntu 20.04 LTS
CUDA	11.7
Python	3.10.4
NumPy	1.17
Pytorch	1.12.1

本章实验所使用的基本实验参数示例如表(6-2),在后续实验中,会改变其中部分参数,这一点在后续章节中详细说明。实验场景设置中,为了保证信号干扰消除算法的性能,基站发天线数量 N_t 最小设置为总用户数量的两倍。本文的根据 3GPP 协议 FR2 频带配置^[47],本实验信号中心频率设置在 4 GHz。通信与感知业务的频带宽度 B_c 与 B_p 设置为相等,均为 10 MHz。通信用户、感知用户均为随机生成在 DFRC 基站周围,具体生成参数如下表所示。

表 6-2 本章基本实验场景参数配置

参数	具体配置
基站天线数 N_t	16
通信用户天线数 N_r	1
通信频带宽度 B_c	10 MHz
感知频带宽度 B_r	10 MHz
信号中心频率 f	4 GHz
通感用户数量 (M, L)	(5, 3)
总发射功率 P^{tot}	$10 \times (M + L)$
噪声功率 P_n	-60 dB
用户方位 θ	Uniform $[0, \pi]$ rad
用户距离 d	Uniform $[1.0, 10.0]$ m

本章所采用的 PPO 算法，其内部深度神经网络有两个，一个是策略网络，另一个是价值函数网络。这两个网络由 3 层全连接网络实现，具体网络结构如表所示。

表 6-3 策略网络结构配置

层类型	维度
Linear	(STATE_DIM, 64)
Tanh	-
Linear	(64, 64)
Tanh	-
Linear	(64, ACTION_DIM)
Sigmoid	-
可训练参数	维度
Logstd	(1, ACTION_DIM)

表 6-4 价值网络结构配置

层类型	维度
Linear	(STATE_DIM, 64)
Tanh	-
Linear	(64, 64)
Tanh	-
Linear	(64, 1)

本文所采用的 PPO 算法存在一系列超参数，如网络学习率、训练步数。具体含义与配置如下：

表 6-5 超参数配置

超参数	具体配置
期望衰减率	0.99
GAE λ	0.95
Minibatch 大小	32
策略更新周期	10
梯度最大值	0.5
学习率	0.0003
训练步数	600000

6.2 算法训练

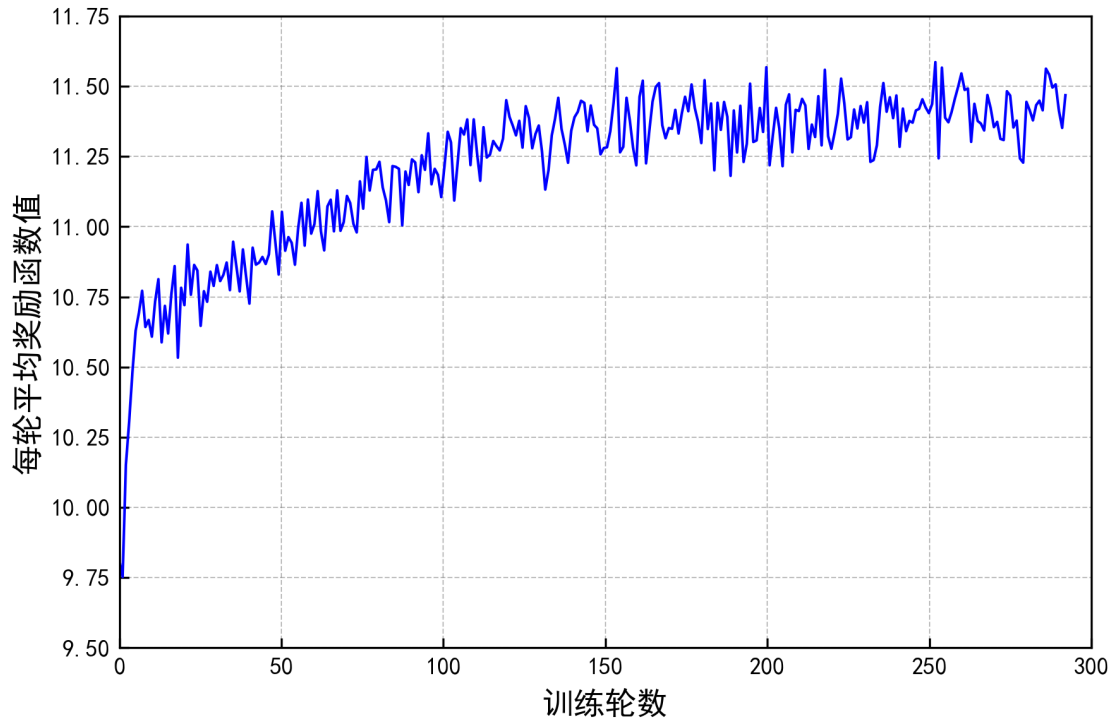


图 6-1 基于 PPO 的功率分配算法训练过程中的平均奖励的收敛性

为了验证基于 PPO 的功率分配算法 (PPO-based Power Allocation, PPO-PA) 在多用户空分复用的通感一体化场景下的效果, 本文首先验证其在训练过程中的收敛性。上图所展示的, 是在基站天线数量为 $N_t = 16$, 通信用户数量为 $M = 5$, 感知用户数量为 $L = 3$ 的情况下, 训练 600000 步过程中, 每个回合的平均奖励的曲线。从上图可以看出, PPO-PA 算法的奖励函数随着训练的进行逐步上升, 并在 150 个回合附近逐渐收敛到最大值, 证明了 PPO-PA 算法的收敛性。

在本章的实验中，奖励函数（目标函数）的参数取值为：

表 6-6 奖励函数配置

超参数	具体配置
ω_1	0.2×10^{-7}
ω_2	10^{-6}
μ	5

ω_1 与 ω_2 的意义是将通信和速率与雷达感知速率和放缩到相近的数量级，并且使得总奖励函数的数量级在 100 以内，这样的目的是防止奖励过大导致网络更新过大，引起梯度爆炸进而使得性能下降。

在下一小节，本文将从通感联合性能、通信性能、感知性能三个维度测评 PPO-PA 算法的性能。本文将在三种场景配置下训练模型，并于相同场景下进行性能测试，与基线算法对比。基线算法采用平均功率分配策略。三种场景设置如下：

表 6-7 场景配置 1

参数	具体配置
N_t	16
M	5
L	3

表 6-8 场景配置 2

参数	具体配置
N_t	16
M	4
L	4

表 6-9 场景配置 3

参数	具体配置
N_t	16
M	3
L	5

6.3 通感联合性能

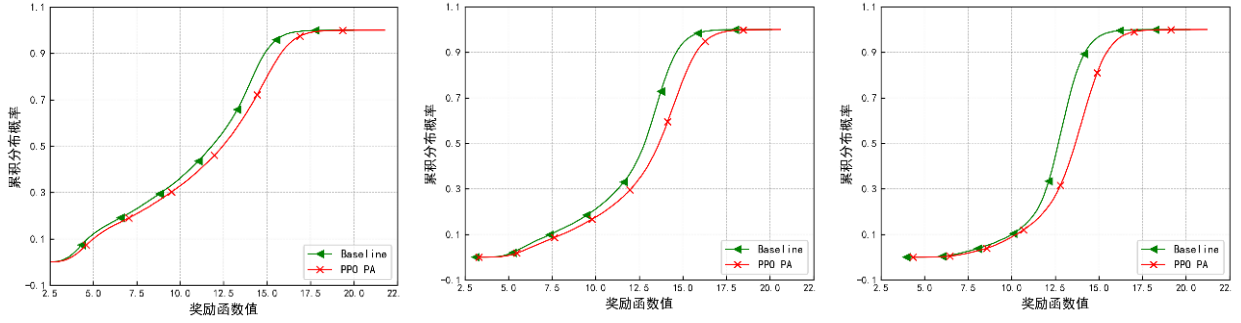


图 6-2 三种场景配置下的奖励累积概率密度曲线

如上图所示，从左到右三张子图依次是：基站天线数量为 $N_t = 16$ ，通信用户数量为 $M = 5$ ，感知用户数量为 $L = 3$ 、基站天线数量为 $N_t = 16$ ，通信用户数量为 $M = 4$ ，感知用户数量为 $L = 4$ 、基站天线数量为 $N_t = 16$ ，通信用户数量为 $M = 3$ ，感知用户数量为 $L = 5$ ，三种场景配置下的实验结果。

本小节将经过 600000 步训练的 PPO-PA 算法的策略网络冻结参数，与基线算法在环境中进行 40000 步的运行，计录奖励函数的值。值得注意的是，本实验的奖励函数即为上一章问题建模中的目标函数加上对违反约束的惩罚项，奖励函数代表着该通感一体化系统的联合性能。

奖励的累积概率密度曲线，反映了 PPO-PA 算法与基线算法在相同的场景配置下得性能，红色的曲线是 PPO-PA 算法，绿色的曲线是基线算法。由于横轴坐标（奖励值）从左向右依次递增，可知累积概率密度曲线越靠右，则代表该算法的奖励偏大的程度越高。因此，如上图，PPO-PA 算法的性能在三种场景下均优于基线算法。

6.4 通信性能

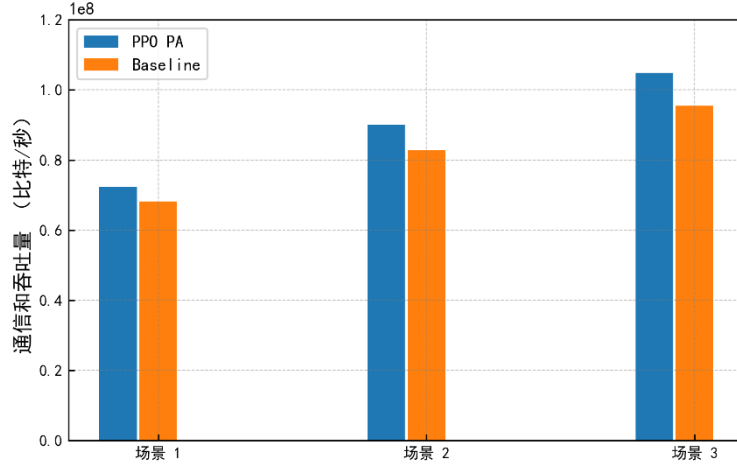


图 6-3 三种场景配置下的通信和吞吐量对比

如上图所示，从左到右三组柱状图依次是：基站天线数量为 $N_t = 16$ ，通信用户数量为 $M = 5$ ，感知用户数量为 $L = 3$ 、基站天线数量为 $N_t = 16$ ，通信用户数量为 $M = 4$ ，感知用户数量为 $L = 4$ 、基站天线数量为 $N_t = 16$ ，通信用户数量为 $M = 3$ ，感知用户数量为 $L = 5$ ，三种场景配置下的实验结果。本小节将经过 600000 步训练的 PPO-PA 算法的策略网络冻结参数，与基线算法在环境中进行 40000 步的运行，统计系统中通信用户的和吞吐量（单位：比特/秒）。

如图上所示，PPO-PA 算法的通信和吞吐量在三种场景设置下均优于基线算法。

6.5 感知性能

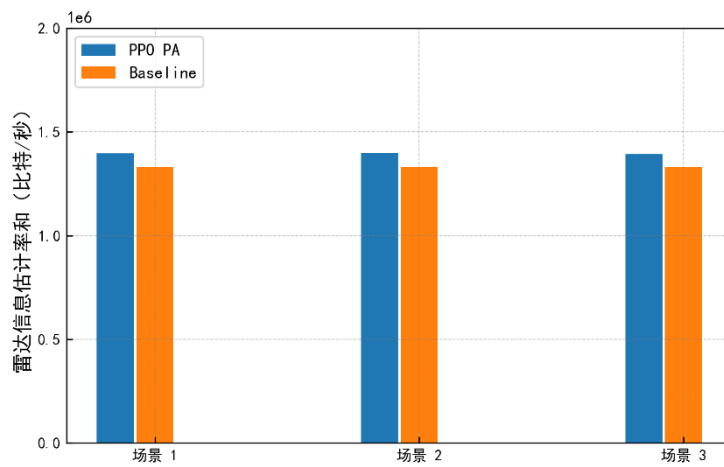


图 6-4 三种场景配置下的雷达感知信息率和对对比

如上图所示,从左到右三组柱状图依次是:基站天线数量为 $N_t=16$,通信用户数量为 $M=5$,感知用户数量为 $L=3$ 、基站天线数量为 $N_t=16$,通信用户数量为 $M=4$,感知用户数量为 $L=4$ 、基站天线数量为 $N_t=16$,通信用户数量为 $M=3$,感知用户数量为 $L=5$,三种场景配置下的实验结果。本小节将经过 600000 步训练的 PPO-PA 算法的策略网络冻结参数,与基线算法在环境中进行 40000 步的运行,统计系统中感知用户的雷达信息估计率和(单位:比特/秒)。

如图上所示,PPO-PA 算法的雷达信息估计率和在三种场景设置下均优于基线算法。

6.6 本章小结

本章对上一章节提出的基于 PPO 的多用户空分复用的资源调度场景下的功率分配算法进行代码实现以及仿真实验,从通感联合性能、感知性能、通信性能等方面验证其性能表现优于基线算法。

7 总结与展望

7.1 工作总结

本文旨在研究通感一体化系统中的资源调度问题。首先,本文针对现有的,针对多用户时分复用场景下的论文进行算法复现及效果评测,对其通信吞吐量、感知准确率、有效吞吐量进行分析评估,分析其优势和劣势,并提出改进方向。

本文设计了一个多用户、多输入多输出的通感一体化网络资源调度系统,采用雷达感知与下行通信相结合,使用空分复用提高通信系统效率和性能,并对场景、信号、信道进行建模,构建数字信号处理流程,以及评价指标的建模。本文基于强化学习算法 PPO,构建了一个基于多用户空分复用场景下的功率分配问题,并提出合理的资源调度系统设计方案,并对基于 PPO 的多用户空分复用的资源调度场景下的功率分配算法进行代码实现及仿真实验,验证其性能表现优于基线算法。

本文设计的基于 PPO 的多用户空分复用场景下的资源调度算法优势如下:

- 本文建模具有普适性,感知与通信目标不耦合,易于推广到更通用的场景;并且,本文构建了较为完整的下行通信、雷达感知的数字信号处理流程。
- 采用多输入多输出技术,有效提高了系统服务于多通信用户、多感知用户的能力。
- 使用雷达信息估计率来表征感知的性能,将感知准确率与频域、带宽、功率建立联系。
- 基于深度强化学习 PPO 的功率分配算法使得本文提出的架构可以支持更为复杂的优化目标求解。

本文也至少存在以下值得改进的方面:

- 没有考虑时域调度的公平性问题,没有考虑当前资源不足以支撑系统同时服务系统所有用户的情况。
- 加权式的感知与通信优化目标是一种非常简单的处理方式,本文仅考虑了使得二者在一定范围内数量级接近的加权参数配置,但没有考虑探究优化目标参数配置对系统性能偏好于感知业务还是通信业务。
- 采取的基线算法过于简单,没有与基于优化的算法或其他智能算法作对比。
- 没有对神经网络运行性能进行评估,且没有使用业内常用的链路级/系统级

信道仿真器进行仿真，缺少系统级仿真与测试。

7.2 未来展望

未来，本文将继续探索通感一体化资源调度系统的优化设计方案，提高其性能和可靠性，并进一步研究其在实际应用中的问题，并提出相应的解决方案，以推动其在实际应用中的广泛应用。主要设想改进方向：

- 增加对用户的时域调度管理。考虑时域调度的公平性问题，并考虑当前资源不足以支撑系统同时服务系统所有用户的情况。
- 进一步探究优化目标函数的设计。探究系统在存在感知或通信业务优先级的情况下的目标函数设计。
- 探究问题规模增加的情况下，资源管理算法需要如何设计以满足网络性能需求。

致谢

时光荏苒。在华科的五年快乐时光已悄然逝去。回首这五年，可谓是“平淡”中有“跌宕起伏”。大二转专业、大三加入种子班、大四决定挑战直博，这些决定是我大学人生的“转折点”，也是一次次成长历练的机遇。这五年中，辗转三个学院，遇到无数老师与同学，是在他们的帮助下，我才能持续前进，不断成长。感谢环境 1807 班、自动化校交 1901 班、种子 1901 班的同学们！感谢华中科技大学 Dian 团队的同学们！感谢辅导员老师教务老师们！感谢交给我们知识的任课老师们！更要特别感谢的，是华中科技大学 Dian 团队的导师们，不论是在学习生活、为人处事，均得到了很多教诲，感谢导师组的辛勤教导！感谢老师们的包容与鼓励！感谢吴远老师、黑晓军老师的教导！感谢 UM 的师兄们的帮助！

很显然，本人文笔不佳，采用了机械式的重复感谢了大家，没能写得感人，但这丝毫不影响我的真诚。

即将离开母校华科，深感不舍。祝愿在华科所有相逢的师生未来更好！

参考文献

- [1] Saad W, Bennis M, Chen M. A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems[J]. IEEE Network, 2020.
- [2] Liu F, Cui Y, Masouros C, 等. Integrated Sensing and Communications: Towards Dual-functional Wireless Networks for 6G and Beyond[J]. arXiv, 2021.
- [3] Ma Y, Zhou G, Wang S. WiFi sensing with channel state information: A survey[J]. ACM Computing Surveys, 2019, 52: 1–36.
- [4] Tan D K P, Du R, Sun Y, 等. Wi-sneeze- sneeze sensing using wi-fi signals[A]. 2020 IEEE globecom workshops (GC wkshps[C]. 2020: 1–6.
- [5] Olivier A, Bielsa G, Tejado I, 等. Lightweight indoor localization for 60-GHz millimeter wave systems[A]. 2016 13th annual IEEE international conference on sensing, communication, and networking (SECON)[C]. 2016: 1–9.
- [6] Shastri A, Valecha N, Bashirov E, 等. A review of millimeter wave device-based localization and device-free sensing technologies and applications[J]. IEEE Communications Surveys & Tutorials, 2022, 24(3): 1708–1749.
- [7] Pin Tan D K, He J, Li Y, 等. Integrated sensing and communication in 6G: Motivations, use cases, requirements, challenges and future directions[A]. 2021 1st IEEE international online symposium on joint communications & sensing (JC&S)[C]. 2021: 1–6.
- [8] Du Z, Liu F, Li Y, 等. Towards ISAC-Empowered vehicular networks: Framework, advances, and opportunities[J]. 2023.
- [9] Liu A, Huang Z, Li M, 等. A Survey on Fundamental Limits of Integrated Sensing and Communication[J]. IEEE Communications Surveys & Tutorials, 2022, 24(2): 994–1034.
- [10] Dong F, Liu F, Cui Y, 等. Sensing as a Service in 6G Perceptive Networks: A Unified Framework for ISAC Resource Allocation[J]. IEEE Transactions on Wireless Communications, 2022: 1–1.
- [11] Yao D, Dai M, Wang T, 等. Intelligent Sensing and Communication assisted Pedestrians Recognition in Vehicular Networks: An Effective Throughput Maximization Approach[A]. IEEE INFOCOM 2022 - IEEE Conference on

- Computer Communications Workshops (INFOCOM WKSHPS)[C]. New York, NY, USA: IEEE, 2022: 1–6.
- [12] Wang Z, Liu Y, Mu X, 等. NOMA Empowered Integrated Sensing and Communication[J]. IEEE Communications Letters, 2022, 26(3): 677–681.
- [13] Chalise B K, Amin M G, Himed B. Performance Tradeoff in a Unified Passive Radar and Communications System[J]. IEEE Signal Processing Letters, 2017, 24(9): 1275–1279.
- [14] del Peral-Rosado J A, Raulefs R, López-Salcedo J A, 等. Survey of cellular mobile radio localization methods: From 1G to 5G[J]. IEEE Communications Surveys & Tutorials, 2018, 20(2): 1124–1148.
- [15] 3GPP. TR 38.855 Study on NR positioning support[J]. 2019.
- [16] Chen K, Zhang D, Yao L, 等. Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities[J]. Acm Computing Surveys, New York, NY, USA: Association for Computing Machinery, 2021, 54(4).
- [17] Lasi H, Fettke P, Kemper H-G, 等. Industry 4.0[J]. Business & Information Systems Engineering, 2014, 6(4): 239–242.
- [18] Moreira A, Prats-Iraola P, Younis M, 等. A tutorial on synthetic aperture radar[J]. IEEE Geoscience and Remote Sensing Magazine, 2013, 1(1): 6–43.
- [19] Sybis M, Vukadinovic V, Rodziewicz M, 等. Communication aspects of a modified cooperative adaptive cruise control algorithm[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(12): 4513–4523.
- [20] Zeng T, Semiari O, Saad W, 等. Joint communication and control for wireless autonomous vehicular platoon systems[J]. IEEE Transactions on Communications, 2019, 67(11): 7907–7922.
- [21] Wang L, Duan Y, Lai Y, 等. V2I-Based platooning design with delay awareness[J]. 2020.
- [22] Milanés V, Villagra J, Godoy J, 等. An intelligent V2I-Based traffic management system[J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(1): 49–58.
- [23] Liu F, Yuan W, Masouros C, 等. Radar-assisted predictive beamforming for vehicular links: Communication served by sensing[J]. IEEE Transactions on Wireless Communications, 2020, 19(11): 7704–7719.

- [24] Yuan W, Liu F, Masouros C, 等. Bayesian predictive beamforming for vehicular networks: A low-overhead joint radar-communication approach[J]. IEEE Transactions on Wireless Communications, 2021, 20(3): 1442–1456.
- [25] Li Y. Deep reinforcement learning: An overview[J]. 2018.
- [26] Sutton R S, Barto A G, others. Introduction to reinforcement learning[M]. MIT press Cambridge, 1998, 135.
- [27] Luong N C, Hoang D T, Gong S, 等. Applications of Deep Reinforcement Learning in Communications and Networking: A Survey[J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3133–3174.
- [28] Bellman R E. Dynamic programming[M]. USA: Dover Publications, Inc., 2003.
- [29] Mao Q, Hu F, Hao Q. Deep learning for intelligent wireless networks: A comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2018, 20(4): 2595–2621.
- [30] Goldsmith A, Jafar S A, Jindal N, 等. Capacity limits of MIMO channels[J]. IEEE Journal on selected areas in Communications, IEEE, 2003, 21(5): 684–702.
- [31] Boyd S P, Vandenberghe L. Convex optimization[M]. Cambridge university press, 2004.
- [32] Bertsekas Dimitrip, 伯塞卡斯, 赵千川, 等. 凸优化理论[M]. 凸优化理论, 2015.
- [33] Schrage L E. Optimization modeling with LINGO[M]. Lindo System, 2006.
- [34] Shaw J. Radiometry and the Friis transmission equation[J]. American Journal of Physics, 2013, 81: 33–37.
- [35] Li J, Stoica P. MIMO radar signal processing[M]. John Wiley & Sons, 2008.
- [36] Nguyen L D, Tuan H D, Duong T Q, 等. Multi-user regularized zero-forcing beamforming[J]. IEEE Transactions on Signal Processing, 2019, 67(11): 2839–2853.
- [37] Richards M A. Fundamentals of radar signal processing[M]. McGraw-Hill Education, 2014.
- [38] Friedlander B. On transmit beamforming for MIMO radar[J]. IEEE Transactions on Aerospace and Electronic Systems, IEEE, 2012, 48(4): 3376–3388.
- [39] Ni Y, Wang Z, Huang Q. Joint transceiver beamforming for multi-target single-user joint radar and communication[J]. IEEE Wireless Communications Letters, 2022, 11(11): 2360–2364.
- [40] Chiriyath A R, Paul B, Jacyna G M, 等. Inner Bounds on Performance of Radar and

- Communications Co-Existence[J]. IEEE Transactions on Signal Processing, 2016, 64(2): 464–474.
- [41] Schulman J, Wolski F, Dhariwal P, 等. Proximal policy optimization algorithms[J]. arXiv preprint arXiv:1707.06347, 2017.
- [42] Schulman J, Levine S, Abbeel P, 等. Trust region policy optimization[A]. International conference on machine learning[C]. PMLR, 2015: 1889–1897.
- [43] Arulkumaran K, Deisenroth M P, Brundage M, 等. Deep reinforcement learning: A brief survey[J]. IEEE Signal Processing Magazine, 2017, 34(6): 26–38.
- [44] Ngo H Q, Larsson E G, Marzetta T L. Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots[A]. 2013 51st Annual Allerton conference on communication, control, and computing (Allerton)[C]. IEEE, 2013: 293–298.
- [45] Huang N-C, Hsieh P-C, Ho K-H, 等. Neural PPO-Clip attains global optimality: A hinge loss perspective[J]. 2022.
- [46] Schulman J, Moritz P, Levine S, 等. High-dimensional continuous control using generalized advantage estimation[J]. 2018.
- [47] TS 38.101-2 NR; User Equipment (UE) radio transmission and reception; Part 2: Range 2 Standalone[J]. 2018.

