

· 专题:5G移动通信基础理论与关键技术 ·

面向5G/B5G通信的智能无线资源管理技术

史清江^{1,4} 洪明毅² 罗智泉^{3,4*}

1. 同济大学软件学院,上海 201804
2. 美国明尼苏达大学电子和计算机工程系,明尼苏达州 55455
3. 香港中文大学(深圳)理工学院,深圳 518172
4. 深圳市大数据研究院,深圳 518172

[摘要] 第五代移动通信(5G)网络能够提供无所不在、万物互联的基础性业务能力。伴随着“新基建”计划的实施,5G将在我国全面部署,而超五代移动通信技术(Beyond 5G, B5G)的研究也正在同步开展。众所周知,高效的无线资源管理是发挥5G/B5G系统效能的关键所在。随着大规模天线、毫米波、网络切片等新技术的出现以及应用场景的多样化和复杂化,无线资源管理问题将达到前所未有的复杂程度。本文在概述5G关键技术的基础上,介绍近几年提出的基于数学优化和人工智能技术的5G/B5G网络资源管理方法,并给出了基于深度学习的无线资源优化和基于大规模优化的网络切片管理两个案例。

[关键词] 5G;无线资源管理;人工智能;优化



罗智泉 加拿大皇家科学院院士,广东省领军人才,麻省理工运筹学博士,北京大学数学学士;现任香港中文大学(深圳)副校长,深圳市大数据研究院院长;发表论文400余篇,获2018年度国际数学优化协会 Paul Y. Tseng 连续优化纪念奖,2015年度 IEEE 信号处理学会最佳杂志论文奖,2010年度 Farkas 奖,2011/2009/2004 年度 IEEE 信号处理学会最佳期刊论文奖,IEEE Fellow, SIAM Fellow。曾任 IEEE 信号处理学会信号处理和通信技术委员会主席,IEEE *Transactions on Signal Processing* 主编。

随着智能终端的普及以及信息化社会的全面发展,人们对于移动通信网络的需求不断增长,全新的应用场景层出不穷,第四代移动通信系统已经无法很好地适应这个万物互联的新时代。从今年开始,第五代(The fifth generation, 5G)移动通信网络将在全球范围内进行部署。相比于第四代移动通信系统,5G的目标是“增强带宽、万物互联”。其典型应用场景包括增强型的移动宽带(Enhanced Mobile Broadband, eMBB)、海量连接的机器通信(massive



史清江 同济大学软件学院教授、博士生导师,深圳市大数据研究院信息系统大数据实验室的访问研究科学家。近年来主要研究领域为智能通信、大数据、优化。主持国家重点研发计划项目、国家自然科学基金青年/面上项目以及华为技术有限公司、中电科等企业院所委托的技术服务项目共计10余项,获得已授权国家发明专利30余项,发表IEEE权威期刊论文50余篇。研究成果获国际会议最佳论文奖(2009),上海市优博(2013),中国通信学会自然科学一等奖(2017),日内瓦国际发明展发明金奖(2018)。曾担任国际信号处理顶级期刊 *IEEE Transactions on Signal Processing* 编委(2017—2019)。

Machine-Type Communications, mMTC),以及高可靠、低时延的联网应用(Ultra-Reliable and Low-Latency Communications, URLLC),并规定了包括频谱效率、峰值速率、时间延迟、可靠性、连接密度及用户体验速率等多个维度的关键技术指标^[1]。围绕上述重要性能指标攻关的5G关键技术包括异构密集网络、大规模天线(Massive MIMO)、毫米波及sub-6G频段通信、新型波形复用与信道编译码、网

收稿日期:2020-03-31;修回日期:2020-04-08

* 通信作者,Email:luozq@cuhk.edu.cn

本文研究受国家自然科学基金项目(61731018,61671411),国家重点研发计划政府间国际合作重点专项项目(2017YFE0119300),广东省领军人才应用基础研究项目(00201501)资助。

络虚拟化与切片技术等。为了突破 5G 关键技术并延伸至 B5G/6G 技术,国内外学术界与工业界已经进行了广泛而深入的研究,取得了很多研究成果。

众所周知,在移动通信系统中,可用的无线资源(包括时、频、空、能、码等多维资源)相对于需求是极其有限的。因此,无线资源管理一直是使能移动通信网络的核心问题。在第五代移动通信系统中,随着大规模天线、毫米波、超密集网络、网络切片等新技术的出现以及应用场景的多样化和复杂化,无线资源管理问题将达到前所未有的复杂程度^[2]。例如,在多用户大规模 MIMO 系统中,随着用户数和天线数的大规模增加,下行用户信道的获取及相应的预编码变得更具挑战,如导频污染问题、高维预编码权重矩阵计算的高复杂度问题等;而在异构超密集网络(Ultra-Dense Network, UDN)中,由于小区尺寸的减少以及各层网络(如宏小区、微小区以及微微小区等)的共存,异层/同层小区间干扰问题以及相邻小区间移动用户的切换问题变得尤为困难;另外,无线网络虚拟化与切片技术虽然能够让 5G“分身有术”支持个性化需求,但也随之带来了新的切片管理与调度问题。这些新挑战和新问题的出现使得 5G 资源管理问题变得愈加复杂,亟需更先进更灵活的无线资源调度方法。

资源调度算法的几个主要关键性能指标包括频谱效率、能量效率、用户 QoS 需求与公平性等。从

系统角度而言,频谱效率和能量效率是最关键的,而从用户角度而言,QoS 需求和公平性更加重要。因此,最佳的调度算法需要考虑这些指标的折衷。基于上述指标,5G 资源管理问题往往被建模成各种无线资源约束下的目标优化问题,包括连续优化问题、离散优化问题及混合规划问题,然后利用先进的(非)凸优化技术设计智能高效的资源调度算法。另一方面,随着人工智能(Artificial Intelligence, AI)科技的进步,尤其是深度学习的发展与应用,通信领域的研究者们也开始关注前沿的人工智能技术(主要包括无监督学习、监督学习、强化学习)^[3],并探索 AI 在 5G/B5G 通信中的应用^[4]。简单讲,利用人工智能技术,可以实现对复杂网络/系统的性能进行准确建模,也可以实现对复杂资源管理问题的高效求解,甚至可以替换部分通信系统模块(如多用户大规模 MIMO 空时处理模块、LDPC/Polar 码信道编译模块等),颠覆通信系统的模块化设计和网络的层次化设计思路,简化网络和系统的实现。图 1 列出了一些用于无线资源管理的典型优化和学习技术。

文章将围绕 5G/B5G 资源管理问题,首先介绍近几年研究者们提出的基于优化技术的 5G/B5G 网络资源管理方法和基于人工智能技术的 5G/B5G 网络资源管理方法。然后,给出基于高效优化技术的网络切片管理和基于深度学习的网络功率/预编码优化两个案例。最后,总结全文并稍稍展望面向 B5G/6G 系统的智能无线资源管理技术。

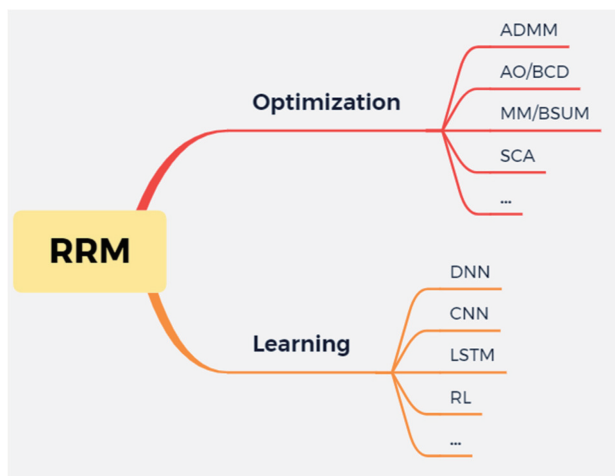


图 1 一些比较流行的用于无线资源管理的优化和学习方法。(RRM: radio resource management; ADMM: alternating directions method of multipliers; AO: alternating optimization; BCD: block coordinate descent; MM: majorization-minorization; BSUM: Block successive Upper-bound minimization; SCA: successive convex approximation; DNN: deep neural network; CNN: convolutional neural network; LSTM: long-short term memory; RL: reinforcement learning.)

1 基于优化的 5G/B5G 无线资源管理技术

数学优化方法已经被广泛应用于通信系统的优化设计中。随着超密集网络、大规模天线、网络切片等新技术的出现,无线资源管理问题变得越来越复杂,规模也變得越来越大,这迫切需要一些高效的大规模优化技术^[5]。本节主要介绍当前一些非常流行的数学优化技术及其在 5G/B5G 通信系统物理层、MAC 层和网络层资源管理问题的应用。

1.1 基于优化的物理层技术

物理层(PHY)是整个 5G 系统设计中最为关键的一环。为了使用更宽的频谱,毫米波段(mmWave, 波段为 3~300 GHz)将被作为一个主要通信波段所采用。毫米波段的波长比较短,所以只需要非常小的天线即可工作,所以很多的天线可以紧密排列在一个很小的区域来一起进行传输,即大规模天线阵列。大规模天线技术和传统的多天线技术(MIMO)有着许多本质上的不同。其中非常重要的一点就是

不可能每一个天线都能分配到一个射频链路(RF chain),否则这样的全数字(FD)设计会导致整个系统价格昂贵。所以在毫米波和大规模天线系统中一个关键技术是所谓的“混合波束形成”技术,即同时使用一个(用移相器网络组成的)射频预编码和一个基带数字预编码来进行上下行的多天线波束形成。Ayach等第一次提出了MIMO的混合波束形成的概念,并且提出这个问题可以看成是一个稀疏重建的问题^[6]。在此基础上,作者提出了一个稀疏优化的数学模型,并且用基追踪(basis pursuit)算法来得到有效的解。Yu等在前文的基础上提出了用交替最小化(Alternating Minimization, AltMin)的算法来交替优化模拟和数字预编码,并且在求解子问题的时候用了流形优化算法(Manifold Optimization)来处理混合预编码的相位所带来的一系列常模约束^[7]。AltMin是一个在非凸优化中常用的方法,这样做的好处是每个孩子问题可能有很好的结构,所以求解会比较简单。以上提到的几项工作都是考虑了点到点的MIMO系统。最近有一些文献讨论了如何更好地在有干扰的多用户系统中设计优化算法。比如^[8]提出了一个基于原始对偶分解(Primal-Dual Decomposition, PDD)的算法。这个算法的好处是将一个有K个用户的混合预编码问题变成了一系列的简单子问题,每个孩子问题可以被分解为K个相对独立的问题来求解,并且可以处理离散相位的要求。总体来说,在混合预编码的设计中,好的优化算法不仅要有很高的计算效率,并且要能处理射频预编码所带来的常模,甚至是离散的约束。

除了预编码之外,大规模天线技术的应用也对其他很多物理层的技术提出了新的挑战。比如在有限的反馈信道带宽下,如何得到最准确的信道信息呢?压缩感知技术在这个方面有着突出的作用。简单来说,虽然需要估计的信道具有很大的维度,但是一般情况下主要多径(Dominant Multipath)的个数很少,所以本质上接收机只需要估计一个低秩或者稀疏的矩阵。利用信道的这个性质,研究者发现,稀疏矩阵优化算法^[9],以及低秩及张量(Tensor)优化算法^[10]在此类问题中起到了至关重要的作用。这些算法的好处是可以利用非常少的反馈信息来精确恢复高维度的信道信息。当然,优化算法在很多其他的物理层的设计上还有很大的应用。比如在大规模天线阵列的设计中,以及在全双工系统波束形成阵列的设计中,基于Majorization-Minimization(MM)的优化算法非常有效。读者如果对MM算法

以及它在信号处理和通信中的应用感兴趣,可以参考Hong的综述文章^[11]。限于篇幅,在这里就不再赘述。

1.2 基于优化的接入层技术

5G网络的一大特点是在一个给定区域同时可能有非常多的设备,它们可能是一般的移动终端,但绝大多数更可能是物联网(IoT)设备。据预测,到2030年,在一个一千平方米的区域中,将会有一千万量级的无线设备,它们随时可能要求接入5G网络,所以在设计接入网络的时候需要考虑的重点包括降低功耗以及大规模接入,这和之前几代的无线通信网络有着很大的不同。在5G接入网的设计中,有很多新的技术的发展都需要设计和运用优化算法,比如非正交多址接入(Non-Orthogonal Multiple Access, NOMA)^[12],大规模接入(Massive Access)等等。

在大规模随机接入的设计问题中,虽然给定区域内的设备的数量很多,但是很多IoT设备并不是每时每刻都需要和网络进行连接。所以基站需要能够有效检测哪些用户正在试图发送信息。对于这个检测问题,一种有效的建模方法是将设备的活动用一个稀疏的“活跃矩阵”来表示,然后从叠加了所有用户的信息中来找出一个最稀疏的矩阵来作为需要估计的活跃矩阵^[13]。在实际系统中,很多情况下用户的信道也可能是未知的,所以需将信道估计和用户检测同时进行。在这样的情况下,前文提过的AltMin算法和原始对偶算法就可以很好地帮助分解问题,提高计算效率。总的来说,各类稀疏优化和低秩优化的算法对于大规模接入非常有效。

另一个非常热门的5G/B5G的接入技术是非正交多址接入(NOMA)。NOMA接入可以让多个用户同时使用同一个载波,以达到可以帮助整个系统显著地提高频谱利用率。Chen等将NOMA的吞吐优化问题建模成一系列的非凸优化问题,分析了最优解的特性,并且定量地说明NOMA系统相对于正交多址接入(Orthogonal Multiple Access, OMA)系统的性能优势^[14]。Tang等讨论了如何结合无线信息和能量传输来提高NOMA系统的能量效率,并提出了一个基于Dinkelbach算法来同时优化信息传输时间和所能接受到的能量^[15]。

1.3 基于优化的网络层技术

5G网络将会不再是只支持个人用户的通信要求,而是一个以多服务为导向的网络(比如车联网、虚拟现实等等)。这将要求整个网络能对于这些不

同的服务进行有效“切片”(Slicing)。网络切片是一种网络功能虚拟化的方法,它允许在一个物理网络的基础上,迅速搭建多个逻辑网络,每个逻辑网络对应一种网络服务^[16]。在这样的网络里面,每个服务是由一系列给定的服务功能组成的,它们叫做服务功能链(Service Function Chain, SFC)。同时,每个网络节点可以提供一些给定的服务功能。所以网络资源管理的任务是尽量满足每种服务的需求(比如用最小的时间完成 SFC),同时保证和每个网络节点不过载。这个问题在计算上的主要难度是它们一般需要离散变量来决定服务、物理节点和用户的一一对应关系。Xu 假设一个服务功能可以在多个节点完成,并且提出了一个具有线性约束的大规模优化问题,然后利用了交替乘法(Alternating Direction Method of Multipliers, ADMM)算法来高效地求解^[17]。Zhang 等提出一个相对复杂和实际的模型,即允许一个 SFC 可以有一系列的功能叠加,并且对于任何一个端对端数据流,每一个的功能必须在一个独立的物理节点上完成^[18]。作者用一个混合整数优化的模型来表示网络切片和服务链选择问题,并且提出了一个基于罚函数的优化方法将整数变量转变为连续变量来求解。总体来说,网络切片问题通常非常复杂,所以在建立数学模型的时候需要很好地平衡模型的精确度,并选择有效的算法来减小计算复杂度。

2 基于 AI 的 5G/B5G 无线资源管理技术

随着人工智能浪潮的再一次兴起,基于机器学习的无线资源管理方法越来越受到学术界和工业界的重视,有望为 5G/B5G 通信系统优化设计提供更智能和高效的解决方案。本节对人工智能技术在物理层、媒体访问控制层和网络层的应用进行系统梳理和回顾。

2.1 基于 AI 的物理层技术

人工智能技术很早就被广泛用于物理层信号识别与调制识别。近年来,随着 5G 最重要的物理层技术——大规模 MIMO 研究的兴起,研究者们提出了许多基于机器学习的大规模 MIMO 系统优化设计方法,包括信号检测、信道估计、功率分配、预编码等,主要目的是降低大规模天线导致的高计算复杂度和反馈开销。

在大规模 MIMO 系统中,特别是当用户规模很大时,庞大的用户数和天线数使得预编码和功率控制等问题的计算复杂度急剧增加,给大规模 MIMO

系统的优化设计带来巨大挑战。Sun 等首次引入深度神经网络解决多小区网络的功率控制问题,验证 DNN 能够有效逼近迭代资源优化算法^[19]。受 Sun 启发,Sanguinetti 等研究了多小区多用户大规模 MIMO 系统的功率控制问题^[20],作者基于用户位置与传播环境特征和网络内干扰大小存在强相关的特性,构建了一个由用户位置到功率分配之间的神经网络映射关系。通过仿真表明,与传统功率分配算法相比,基于 DNN 的方法可以大大降低功率分配的计算复杂度。同时,由于无需计算信道的统计特性,也大大方便了计算。Xia 等拓展了上述功率分配工作,研究了基于 DNN 的多用户 MISO 下行系统预编码优化设计,利用预编码最优结构这一先验信息,降低神经网络的输出维度,从而增强神经网络拟合的容易程度^[21]。

部署大规模 MIMO 系统需要解决的另一大难题是如何进行高效的信道状态估计和反馈。由于经典的信道模型在大规模 MIMO 场景下不再准确,传统的基于信道建模的估计方法难以适用。这激发了一系列数据驱动的解决方案的研究。例如,针对大规模 MIMO 系统,He 等提出了一种基于学习去噪的 AMP (Learned Denoising-Based AMP, LDAMP)信道估计方法^[22]。该方法巧妙地将信道矩阵看作二维自然图像,使得在图像处理上获得巨大成功的卷积神经网络得以引入,并与传统的 AMP 算法相结合以更好地捕捉大规模 MIMO 的信道特性。随后,受压缩感知算法启发,Wen 等提出了一种新型的称之为 CsiNet 的神经网络以进一步减少大规模 MIMO 系统的信道反馈开销^[23]。CsiNet 的网络结构是通过模拟压缩感知结构得到的,其中包含一个卷积神经网络以对信道矩阵进行特征提取和一个自动编码器以对信道矩阵进行压缩和重建。在此基础上,Wang 等将 CsiNet 与一个称为 LSTM 的递归神经网络进行级联,以进一步挖掘大规模 MIMO 信道的时延相关性,从而提高估计精确度^[24]。

2.2 基于 AI 的接入层技术

对于日趋复杂的 5G 网络环境,设计符合不同网络环境和各种应用需求的 MAC 协议是一项大挑战,尤其是当无法有效建模这些网络环境和应用需求的时候。机器学习允许人们从大量数据中学习到网络环境的相关特征及变化,能有效地应对这一挑战。

在 5G 时代,无线设备的快速增多,对频谱资源的需求急剧增加,这时就需要动态频谱共享。为了

以更有效的方式共享可用频谱,研究者在预测频谱可用性和提高网络的整体性能方面进行了各种尝试。然而,传统的基于监督学习的频谱预测方法需要收集大量的标记数据来进行有效的模型训练,这一过程十分耗时。另外,在一些快速变化的环境中,很难针对每种环境状态都收集大量数据来确保模型训练的准确性。为此,深度强化学习成为了目前比较热门的替代方案。与监督学习不同,深度强化学习不通过标记数据进行学习而通过智能实体与环境的直接交互来捕捉环境变化的规律并学习最优的策略。目前已有研究将深度强化学习应用在动态频谱检测和 MAC 协议选择上。例如,针对多时隙网络中时隙共享的问题,Yu 等提出一种基于残差网络的深度强化学习解决方案来最大化整个网络的吞吐量^[25]。该方案能通过一系列的观察和操作来预测频谱空闲并学习一种最佳的时隙共享策略,而无需知道其他共存网络 MAC 协议的运行机制。类似地,针对无线传感网络,Wang 等提出了一种基于深度 Q 网络的多址接入解决方案来预测和选择最优的接入信道,以最大化平均网络吞吐量^[26]。

2.3 基于 AI 的网络层技术

在 5G 时代,为了进一步优化网络的整体性能,运营商需要对网络性能进行预测,以便让网络参数设置实时适配不断变化的网络环境和应用的服务质量需求。在众多模型中,基于机器学习的预测模型由于其通用性和灵活性是目前研究的热点。例如,Pierucci 将基于神经网络的预测模型应用于预测蜂窝网络中的用户体验^[27]。该模型充分挖掘了多种输入参数(包括平均用户吞吐量、单元中有效用户数、每个用户的平均数据量和通道质量指示器等)与用户体验之间的相关性并展现出很高的预测精度。

人工智能技术也引入到了网络切片管理问题。例如,Bega 等采用了基于深度卷积神经网络的预测模型,该模型利用在特定网络切片的基站上观察到的流量热力分布图,来预测未来相关服务的流量需求^[28]。为了进一步提高流量预测的准确性,Wang 等提出了一种称为 LSTM 的循环神经网络模型以挖掘流量的时间相关性,并将该模型与通过自动编码器提取的流量空间特征相结合进行流量预测^[29]。该方案在基于真实数据集的实验展现出远超传统方法的性能表现。最近,为了更高效地挖掘流量的时间与空间相关性,Fang 等提出了一种结合图卷积神经网络和 LSTM 的新型深度学习架构来预测移动流量,并取得了更显著的性能提升^[30]。

3 典型案例

本节将具体介绍两个典型范例:基于机器学习的无线资源管理和基于大规模优化的网络切片管理。

3.1 基于机器学习的无线资源管理

几十年来,迭代数值优化在解决无线通信、信号处理任务中起到了关键作用。在迭代算法中,优化问题的参数是迭代算法的输入,而运行许多次迭代后的结果是迭代算法的输出。当优化问题的规模较大,特别是涉及到矩阵求逆等计算,迭代算法的计算复杂度就会很高,往往无法满足通信系统中的实时性任务要求。而 DNN 具有强大的函数逼近能力,它能够有效逼近迭代算法的性能并大大降低计算复杂度。

下面以 Sun 等的研究^[19]为例,阐述 DNN 用于无线资源管理的基本思路。如图 2 所示,基本步骤可归纳如下:

Step 1:对于一类无线资源管理问题,设计迭代资源优化算法;

Step 2:利用迭代算法的特点设计深度神经网络。一般来讲,将迭代算法的输入参数作为神经网络的输入,迭代算法的输出(也就是资源管理策略)作为神经网络的输出;

Step 3:对于每个问题实例,利用迭代资源优化算法计算得到最优的资源管理策略,形成训练样本集;

Step 4:选择损失函数然后训练神经网络得到网络模型;

Step 5:对于新的问题实例,利用神经网络模型计算资源管理策略。

利用上述设计思路,可以求解几乎所有无线资源优化问题(如复杂的随机优化,整数规划等),可以有效提升资源分配策略的计算速度。在选择神经网络类型时,除了一般的前馈神经网络,也可以考虑卷积神经网络(当输入是信道矩阵时)和图神经网络,后者已被证明能够有效求解整数规划问题。而在设计神经网络时,以下两点值得研究者们特别注意:(1)一般无线资源优化问题的目标函数是系统效用,如系统频效、能效等。因此,对于面向无线资源管理的神经网络,除了可以选均方误差函数(对应监督学习)作为神经网络的损失函数,也可以直接选系统效用函数(对应无监督学习)作为神经网络的损失函数;(2)可以利用无线资源优化问题的最优解结构,将算法的先验信息融入到神经网络设计中,简化神经网络的输入输出设计,在加速神经网络训练速度的同时可以有效增强神经网络逼近迭代算法的能力。

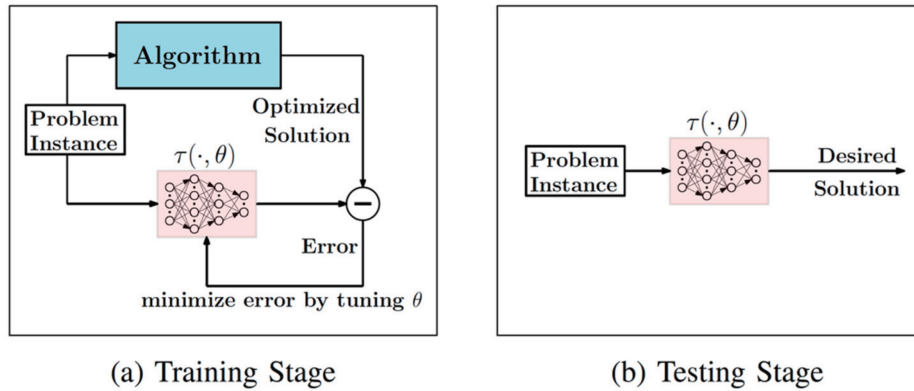


图 2 算法驱动深度学习无线资源管理技术

3.2 基于大规模优化的网络切片管理

如第二节中所述，网络切片优化的主要难度是这个问题一般需要引入离散变量来决定服务、物理节点和用户的一一对应关系。下面以 Zhang 等的研究^[18]为例，阐述以下建模和求解的基本思路。

考虑一个有线网络，它可以用一个图 $G = (V, L)$ 来表示，其中 $V = \{v_i\}$ 是一组网络节点， $L = \{l_{ij}\}$ 是一组有向链路。用 V_f 来表示一组节点，他们可以提供功能 f 。一共有 K 个不同的数据流，每个数据流 k 需要一系列服务，用一个有序集合 $F(k) = (f_1^k \rightarrow f_2^k \rightarrow f_3^k \dots \rightarrow f_n^k)$ 来表示。用 $r_{ij}(k)$ 来表示数据流 k 在链路 (i, j) 上的数据。为了更好地建立服务功能和物理节点之间的关系，需要定义一系列的二元变量。用二元变量 $x_{if}(k) \in \{0, 1\}$ 来表示节点 i 是不是为数据流 k 提供了服务 f 。对于数据流 k 来说，一个重要的约束条件是数据流 k 所需要的服务按服务规定的顺序 $(f_1^k \rightarrow f_2^k \rightarrow f_3^k \dots \rightarrow f_n^k)$ 来逐一满足。所以，优化问题可以概括地写成^①

$$\begin{aligned} & \min \sum_{k, (i, j)} r_{ij}(k) \\ & s. t. \quad [\text{每个链路容量不超负荷}] \\ & \quad \quad [\text{每个数据流的每个服务被一个节点满足}] \\ & \quad \quad [\text{每个链路容量不超负荷}] \\ & \quad \quad [\text{每个节点对每个流只服务一次}] \\ & \quad \quad [\text{每个节点的计算容量不超负荷}] \end{aligned}$$

虽然没有仔细写出每个约束的数学表达式，但是相信读者可以发现这是一个非常复杂的混合整数优化问题。它不仅有着非常多的约束和变量，并且很多变量是整数变量，比如 $x_{if}(k) \in \{0, 1\}$ 。

那么如何解决这个问题呢？Zhang 等提出了一个行之有效的“松弛—惩罚”方法。它的主要思想是：(1) 将每个 $x_{if}(k)$ 的值域从 $\{0, 1\}$ 松弛到 $[0, 1]$ ，

这样的话整个问题就变成了一个线性规划问题；(2) 为了保证松弛的二元变量还尽量维持在 0 或者 1，在目标函数上加行了惩罚项 (Penalty Term)，即一个 p -范数有关的一个非凸函数，来惩罚非整数的 $x_{if}(k)$ 。然后采用分块连续上界最小化 (Block Successive Upper-Bound Minimization, BSUM) 方法来求解松弛和惩罚了之后的问题。整个算法被称作 Penalty-BSUM (PBSUM)，它不仅能够非常有效的求解一个大规模的网络资源优化问题，更难能可贵的是，它所得出来的解质量非常高。为了展示 P-BSUM 算法在实际中的应用，考虑一个有 112 个节点 400 条链路的网络 (见图 3)。在图 4 中可以看到 PBSUM 算法和它的一个衍生算法 PBSUM-R 的表现非常好，在 50 次仿真中，每次这些算法都可以得到接近全局最优的解 (ξ 的值非常接近 1)。

总体来说，网络切片设计是具有很大难度的工程问题。精确的数学模型通常同时涉及许多离散和连续变量。在上例中讨论的“松弛—惩罚”策略是一个有效减小计算复杂度的方法，可以非常容易推广到其它混合整数规划型无线资源管理问题。

4 结 语

随着新基建计划的实施，5G 网络将成为支撑全行业、全社会的基础性设施。为了更好地发挥 5G/B5G 网络的效能，高效的无线资源管理是关键所在。然而，随着新技术和新应用场景的出现，无线资源管理问题变得愈加复杂。为了应对挑战，传统的基于集中式优化的无线资源管理方法需要朝着分布式优化方向发展，如采用大规模优化解决超密集网络资源优化、大规模天线系统优化设计等。此外，随着

① 限于篇幅关系，本文只写出一个最概括的优化模型，把细节都省略了，有兴趣的读者可以参考[18]。

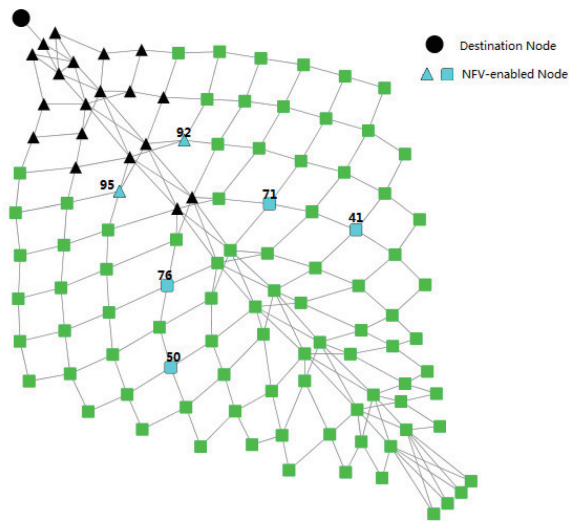


图3 一个有着112个节点,440条链路6个可以提供服务的节点, $K=20$ 个数据流,每一个数据流的数据率是在 $[1,5]$ 之间;每条链路的容量是在 $[5,50]$ 之间,每个节点的计算容量是16;黑的圆圈表示目标节点,蓝色三角形和方块表示可以提供服务的节点。绿色和黑色三角形方块表示一般节点,黑色三角形节点不能作为数据流的起始节点。

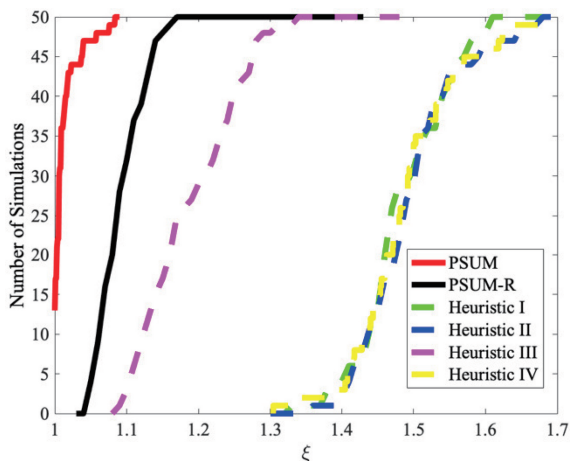


图4 不同版本的P-BSUM和一些现有的启发式算法的比较。一共产生了50个不同的网络。对于每个网络用不同算法算出系统吞吐,然后和最优解进行比较; ξ 是最优解和优化解的比值,越接近1代表着算法的效果越好。

AI技术的迅猛发展, AI与通信融合为超越传统的基于模型的无线资源优化方法的设计理念与性能提供了潜在的可能性。特别地,数据和模型双驱动的无线资源管理技术有望以更灵活通用的形式适应变幻莫测的无线环境,并以更智能高效的方式对无线资源进行动态规划和分配。

无论基于优化还是AI的无线资源管理方法,都已有大量的研究文献可供借鉴。本文并非试图全面地总结该领域的已有研究成果,而是希望梳理出对未来无线网络(如6G)资源管理有潜在价值的典型

优化技术和AI技术,期待这些技术在不远的将来发挥重要作用。

致谢 感谢深圳市大数据研究院朱光旭博士和同济大学博士生周凌云、硕士生颜子彦对本文准备阶段工作的支持。

参 考 文 献

- [1] 尤肖虎, 潘志文, 高西奇, 等. 5G移动通信发展趋势与若干关键技术. 中国科学:信息科学, 2014, 44(5):551—563.
- [2] Calabrese FD, Wang L, Ghadimi E, et al. Learning radio resource management in RANs: framework, opportunities, and challenges. IEEE Communications Magazine, 2018, 56(9): 138—145.
- [3] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553): 436—444.
- [4] 尤肖虎, 张川, 谈晓思, 等. 基于AI的5G技术——研究方向与范例. 中国科学:信息科学, 2018, 48(12):5—18.
- [5] Baligh H, Hong MY, Liao WC, et al. Cross-layer provision of future cellular networks: a WMMSE-based approach. IEEE Signal Processing Magazine, 2014, 31(6), 56—68
- [6] Ayach OEI, Rajagopal S, Abu-Surra S. Spatially sparse precoding in millimeter wave MIMO systems. IEEE Transactions on Wireless Communications, 2013, 13(3):1499—1513.
- [7] Yu XH, Shen JC, Zhang J. Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems. IEEE Journal of Selected Topics in Signal Processing, 2016, 10(3), 485—500.
- [8] Shi QJ, Hong MY. Spectral efficiency optimization for millimeter wave multiuser MIMO systems. IEEE Journal of Selected Topics in Signal Processing, 2018, 12(3): 455—468.
- [9] Bajwa WU, Haupt J, Sayeed AM, et al. Compressed channel sensing: a new approach to estimating sparse multipath channels. Proceedings of the IEEE, 2010, 98(6): 1058—1076.
- [10] Qian C, Fu X, Sidiropoulos ND. Tensor-based channel estimation for dual-polarized massive MIMO systems. IEEE Transactions on Signal Processing, 2018, 66(24): 6390—6403.
- [11] Hong MY, Razaviyayn M, Luo ZQ. A unified algorithmic framework for block-structured optimization involving big data: with applications in machine learning and signal processing. IEEE Signal Processing Magazine, 2015, 33(1):57—77.
- [12] Liu YW, Qin ZJ, El Kashlan M, et al. Nonorthogonal multiple access for 5G and beyond. Proceedings of the IEEE, 2017, 105(12):2347—2381.
- [13] Chen XM, Ng DWK, Yu W, et al. Massive access for 5G and beyond. arXiv preprint arXiv:2002.03491.
- [14] Chen ZY, Ding ZG, Dai XC, et al. An optimization perspective of the superiority of 5G NOMA compared to conventional OMA. IEEE Transactions on Signal Process, 2017, 65(19):5191—5202.
- [15] Tang J, Luo JC, Liu MQ, et al. Energy efficiency optimization for NOMA with SWIPT. IEEE Journal of Selected Topics in Signal Processing, 2019, 13(3):452—466.

- [16] Foukas X, Patounas G, Elmokashfi A, et al. Network slicing in 5G: survey and challenges. *IEEE Communications Magazine*, 2017, 55(5):94–100.
- [17] Xu H, Li BC. Joint request mapping and response routing for geodistributed cloud services. *Proceeding IEEE INFOCOM*, 2013, 854–862.
- [18] Zhang N, Liu YF, Farmanbar H, et al. Network slicing for service-oriented networks under resource constraints. *IEEE Journal on Selected Areas in Communications*, 2017, 35(11):2512–2521.
- [19] Sun HR, Chen XY, Shi QJ, et al. Learning to optimize: training deep neural networks for interference management. *IEEE Transactions on Signal Processing*, 2018, 66(20): 5438–5453.
- [20] Sanguinetti L, Zappone A, Debbah M. Deep learning power allocation in massive MIMO. *Signals Systems and Computers 2018 52nd Asilomar Conference on*, 2018, 1257–1261.
- [21] Xia WC, Zheng G, Zhu YX, et al. A deep learning framework for optimization of MISO downlink beamforming. *IEEE Transactions on Communications*, 2020, 68(3): 1866–1880.
- [22] He HT, Wen CK, Jin S, et al. Deep learning-based channel estimation for beam space mmWave massive MIMO systems. *IEEE Wireless Communications Letter*, 2018, 7(5): 852–855.
- [23] Wen CK, Shih WT, Jin S. Deep learning for massive MIMO CSI feedback. *IEEE Wireless Communications Letter*, 2018, 7(5): 748–751.
- [24] Wang TQ, Wen CK, Jin S, et al. Deep learning-based csi feedback approach for time-varying massive MIMO channels. *IEEE Wireless Communications Letter*, 2019, 8(2): 416–419.
- [25] Yu YD, Wang TT, Liew SC. Deep-reinforcement learning multiple access for heterogeneous wireless networks. *IEEE Journal on Selected Areas in Communications*, 2019, 37(6): 1277–1290.
- [26] Wang SX, Liu HP, Gomes PH, et al. Deep reinforcement learning for dynamic multichannel access in wireless networks. *IEEE Transactions on Cognitive Communications and Networking*, 2018, 4(2): 257–265.
- [27] Pierucci L, Micheli D. A neural network for quality of experience estimation in mobile communications. *IEEE Multi Media*, 2016, 23(4): 42–49.
- [28] Bega D, Gramaglia M, Fiore M, et al. Cognitive network management in sliced 5G networks with deep learning. *Proceeding IEEE Conference on Computer Communications*, 2019, 280–288.
- [29] Wang J, Tang J, Xu ZY, et al. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. *Proceeding IEEE Conference on Computer Communications*, 2017, 1323–1331.
- [30] Fang LY, Cheng X, Wang HN, et al. Mobile demand forecasting via deep graph-sequence spatiotemporal modeling in cellular networks. *IEEE Internet of Things Journal*, 2018, 5(4)(SD): 3091–3101.

Intelligent Radio Resource Management for 5G/B5G Communications

Shi Qingjiang^{1,4} Hong Mingyi² Luo Zhiqian^{3,4*}

1. *School of Software Engineering, Tongji University, Shanghai 201804*

2. *Dept. of ECE, University of Minnesota, MN 55455*

3. *School of Science and Engineering, The Chinese University of Hongkong (Shenzhen), Shenzhen 518172*

4. *Shenzhen Research Institute of Big Data, Shenzhen 518172*

Abstract The 5G wireless system is capable of providing ubiquitous connections, and it is the foundation for realizing the full potential of Internet of Things (IoT). Starting 2020, the 5G system will be deployed nationwide, and the Beyond 5G (B5G) technologies will undergo extensive research. However, the new technologies to be introduced in different layers of the 5G systems, such as massive access, millimeter wave, and network slicing, pose significant design challenges for effective wireless resource allocation. In this work, we discuss how tools from mathematical optimization and machine learning can facilitate the design of resource allocation tasks arising from PHY, MAC and network layers of 5G systems. We start the paper by giving an overview of recent advances, followed by two specific examples of how optimization and machine learning can be used to design highly efficient network slicing and beam forming algorithms.

Keywords 5G; radio resource management; AI; optimization

(责任编辑 姜钧译 吴妹)

* Corresponding Author, Email: luozq@cuhk.edu.cn