

南京信息工程大学

本科生毕业论文 (设计)



题 目 公共交通调度策略的最优化方法
 ——以南京信息工程大学校园为例

学生姓名 陈冰

学 号 201933070085

学 院 应用技术学院

专 业 软件工程

指导教师 马杰良

二〇二三年二月一日

声 明

本人郑重声明：

- 1、 以“求实、创新”的科学精神从事科学研究工作。
- 2、 本论文中除引文外，所有测试、数据和相关材料均为真实有效的。
- 3、 本论文是我个人在指导教师的指导下进行的研究工作和取得的研究成果，请勿用于非法用途。
- 4、 本论文中除引文和致谢的内容外，并未抄袭其他人或其他机构已经发表或撰写过的研究成果。
- 5、 关于其他同志对本研究所做的贡献均已在论文中作了声明并表示了谢意。

作者签名：

日期： 年 月 日

目 录

1	绪言.....	1
1.1	研究背景和意义	1
1.2	研究现状.....	1
1.3	研究内容.....	4
1.4	论文组织结构	4
2	研究相关理论.....	5
2.1	排队论	5
	参考文献	6

南京信息工程大学本科生毕业论文 L^AT_EX 模板 V2021.6¹

Bruce Y.P. Lee²、LiR³、John D⁴、B. Shen⁵

南京信息工程大学 L^AT_EX 学院，江苏 南京 210044

摘要：这是一份南京信息工程大学本科生毕业论文 L^AT_EX 模板。友善提醒：本文档是非官方版，属个人兴趣产物。

关键词：模板；南信大；毕业论文；

¹本模板制作时间：2014 年 5 月，最后修订于 2021 年 6 月

²E-mail: yupenglee119@gmail.com

³第二版修改者，E-mail: stuliren@outlook.com

⁴2021.6 版修改者，E-mail: mailto:work.temp.place@outlook.com

⁵2022 版修改者，E-mail: nj_bwshen@outlook.com

L^AT_EX Template for Undergraduate Thesis of Nanjing University of Information Science and Technology

Bruce Y.P. Lee、LiR、John D、B. Shen

School of L^AT_EX, NUIST, Nanjing 210044, China

Abstract: This is a L^AT_EX template for the Undergraduate thesis of Nanjing University of Information Science and Technology. Caution: due to personal interest, not an official template.

Key words: template; NUIST; thesis;

1 绪言

1.1 研究背景和意义

公共交通，泛指所有向大众开放、并提供运输服务的运输系统，该系统通常按照固定的时刻表进行资源调度管理，在既定的线路上运行，每次向乘客收取一定的费用。广义而言，公共交通包括民航、铁路、公路、水运等交通方式。公交系统作为城市公交网络的重要组成部分，其发展水平直接影响了市民出行的便捷性和幸福感，对建设和谐社会及和谐城市有着重大意义。公交系统与居民的日常出行息息相关，是居民日常出行的主要选择。以南京市公共交通系统为例，2022 年第一季度，全市营业性客车达 6353 辆，客位 279619 座，全市客运班车总数达 879 辆、45428 客位，旅游及包车 5474 辆、240286 客位。普通干线公路日均交通量 29.52 万辆^[1]。但是城市人口的快速增长造成了交通拥堵，同时交通拥堵也阻碍了城市经济的发展。根据高德地图发布的《2022 年度中国主要城市交通分析报告^[2]》，2019 至 2022 年度大中型城市的整体候车时长同比呈上升趋势，尤其是受发车频率影响的候车时长上升明显；主要城市高峰平均交通拥堵指数达到 1.704（在通行一小时的行程中实际花费为 1.704 小时）。根据百度地图发布的《2022 年度中国城市交通报告^[3]》，我国主要城市公交候车时间为 8.97~17.34 分钟，且候车时间越长的城市受发车频率和交通扰动影响也越大。倡导公交出行是解决交通拥堵的一个主流方案，城市公交系统对节能减排，建设绿色城市也有着重要意义，公交系统依赖城市道路网络，且大多为电力驱动，更多的载客量下产生的碳排放更少，且良好的站点设置与调度方式可以让公交系统与共享单车、轨道交通等其他出行方式结合产生完美的绿色出行生态系统。如何构建一个高效稳定的交通系统，提高乘客的出行满意度，以吸引更多乘客选择公共交通出行，对推动绿色出行和社会发展有着重大意义。

交通调度优化是一项十分复杂的系统性工程，多种因素相互合作用影响着整个系统。随着大数据技术的发展，各行各业每天产生海量的数据，越来越多的特征被加入模型中，海量数据的支持让机器学习在各行各业崭露头角。使用机器学习与数据挖掘算法，通过对道路、天气、乘客等数据的挖掘分析，可以更准确的建立乘客出行模式的模型。在建立了乘客出行模式的模型后，可以对客流量进行更好的预测。传统的运筹学方法对公交调度这类有约束资源分配问题下有良好的表现，但是在大数据时代难以发挥作用，将机器学习与运筹学结合，使用机器学习算法更好的对不同场景下的乘客出行喜好进行分析、对客流量进行预测。通过所得的数据建立相应问题的规划模型并求解，更好的解决相应交通规划问题。

本文以南信大校园公交系统为例，探究在大数据时代下，如何结合传统运筹优化算法与机器学习的优势，对公交系统进行优化。现有运营方案下，校园公交人流量高峰时期常采用多车次连续发车的方法，由于公交线路站点大多均匀坐落在校园主干道上，高峰时期车辆行进速度也会受到主干道交通环境影响，导致部分站点排队时间较长；非高峰时期，为考虑运营成本，公交采用固定弹性时刻表方式进行调度，现有的调度策略存在滞后性。通过对客流大数据进行挖掘分析，运用机器学习算法对客流量模式进行识别，通过该模式建立数学模型，使用运筹学方法从成本、效益、乘客舒适度等多个角度出发，对公交站定址、调度策略等进行优化，最终解决现有策略下校园公共交通系统中的缺陷。

1.2 研究现状

排队理论中常用分布函数来描述客流，近年来基于机器学习的客流预测方法受到了广泛的关注。在机器学习任务中，诸如预测上证指数、商场每日人流量、交通客流这类一组按照时

间发生先后顺序进行排列的数据点序列，被称为时间序列预测。时序预测中一种常用的算法是 LSTM。LSTM，全称 Long Short Term Memory (长短期记忆) 是一种特殊的递归神经网络。2022 年，Wenhua Jiang 等人提出一种基于深度学习的短期 OD（始发地-目的地）客流预测方法^[4]，短期 OD 流量预测的一个关键挑战是由于出行没有在一定的时间间隔内完成，OD 流量信息的部分可观测性。该方法开发了一种用于城市轨道交通系统 OD 流量预测的新型深度学习架构，并研究了数据表示和处理部分信息的各种机制。深度学习框架由三个主要部分组成，包括多个 LSTM 网络，该网络具有捕获短期/长期时间依赖性的注意机制，用于时空相关性的时间移位图矩阵，以及用于部分 OD 流观测的重建机制。结果表明，所提出的模型具有较高的精度和鲁棒性，以及 OD 流量信息的局部观测对提高预测性能的重要性。在数据表示方面，预测 OD 流量偏差的效果始终优于直接预测 OD 流量。

LSTM 与一般的前馈神经网络不同，LSTM 可以利用时间序列对输入进行分析；简而言之，当使用前馈神经网络时，神经网络会认为 t 时刻输入的内容与 $t + 1$ 时刻输入的内容完全无关，对于许多情况，例如图片分类识别，这是毫无问题的，可是对于一些情景，例如自然语言处理 (NLP, Natural Language Processing) 或者我们需要分析类似于连拍照片这样的数据时，合理运用 t 或之前的输入来处理 $t + n$ 时刻显然可以更加合理的运用输入的信息。在 2017 年提出的基于 self-attention 机制的 Transformer 模型^[5]最开始在自然语言处理任务中有着不俗的表现，该模型的思想后被广泛应用于机器学习的各个任务分支中，最近的一些研究表明，在对于能源、交通、天气预测等时序预测任务中，Transformer 同样有着优于 LSTM 和 RNN 的表现，尤其是包含双向时空适应（空间自适应和时间自适应）结构的 Transformer^[6-7]。在交通流量预测问题中，现有方法多侧重于时空依赖建模，而忽略了交通预测问题的两个内在特性。不同预测任务的复杂性在不同的空间（如郊区与市中心）和时间（如高峰时段与非高峰时段）上分布不均匀。其次，对过去交通状况的回忆有利于对未来交通状况的预测。基于以上两个特性，一个双向时空自适应 Transformer (Bi-STAT) 可以用于准确的交通流预测。Bi-STAT 采用编码器-解码器框架，其中均含有一个空间自适应和时间自适应的 Transformer 结构。受第一个性质的启发，每个 Transformer 都根据任务的复杂性动态地处理流量流，通过一种新的动态停止模块 (DHM) 的循环机制来实现这一点。每个 Transformer 使用共享参数进行迭代计算，直到 DHM 发出停止信号。受第二个特性启发，Bi-STAT 使用一个解码器实现现在-过去的学习任务，另一个解码器实现现在-未来的预测任务。学习任务提供补充信息协助预测任务，以便更好地泛化。大量实验证明了 Bi-STAT 每个模块的有效性以及该模型的优越预测性能。

常规情况下，公共交通调度采用有规则定期时刻表的方式进行调度，该方法下会产生一个非线性混合整数模型。2004 年，Alessandro Chierici 等人扩展了此种普遍被采用方法^[8]，新方法考虑时间表的质量与交通系统对其他可替代的运输方式吸引的乘客之间的相互影响。在放宽完整性约束后，所得的非线性混合整数模型仍然是非凸的。通过基于外部近似的分支定界算法和利用两个子模型的分解和相互更新的启发式算法来解决它。随着机器学习近十年的飞速发展带动了各行各业对于数据驱动的研究热潮，运筹学也不例外，2020 年 Google 与 Deepmind 提出一种基于深度神经网络提升传统混合整数规划方法的性能^[9]。传统的混合整数规划中通常是基于分支定界法 (branch and bound)，通过不断求解连续凸松弛问题找出最优解，但是理论上需要求解的松弛问题个数随着整数变量维度上升呈指数倍数增长。因此，实际中会添加许多加速方法去减少需要求解的松弛问题数量，其中有很多加速方法的效果取决于混合整数规划的问题结构或是当前分支定界已有的信息，但是很多时候由于模型过于复杂，可能无法直接提取出这类信息，而这正好是深度学习方法的优点，它能基于充足的数据通过黑盒优化提供一个预测模型，捕捉一些难以解释的有效信息。核心是使用神经网络提升传统混合整数

规划中两种常见的启发式方法的性能：下潜（Diving）搜索和分支（Branching）选取，其主要思想还是利用神经网络基于数据的预测能力对某一类特定问题求解实现加速。

在求解需求为动态的一类运输问题时，常使用动态规划的方法来切结列车调度问题。2020年，Renming Liu 等，针对一类时间依赖需求的地铁列车节能调度问题，提出了列车交通模型^[10]，包含列车车头时距、列车载客量和地铁沿线能耗演化的动态方程，制定了非线性动态规划（DP）问题以生成近似最优时间表，进而实现列车利用率、乘客等待时间、服务水平和能源消耗之间的协调。为了克服维度灾难问题，构建了一个近似动态规划（ADP）框架，引入了状态、策略、状态转换和奖励函数的概念。通过数值实验验证了所提模型和算法的有效性，并与遗传算法和差分进化算法进行了比较，该算法能够在较短的时间内收敛到一个较好的解。Pengli Mo 等人的研究^[11]针对城市轨道交通线路，结合动态客流需求特征，分析列车运行图与列车运用计划的内在联系，建立列车运行图与车底运用计划协同节能优化模型；该模型通过提高牵引-制动重叠时间和线上折返次数，以实现城轨运营中列车运行图与车底运用计划的协同优化，基于北京地铁亦庄线的运行数据验证了将运行图和车底运用计划分开考虑可能会导致车底的不充分利用和接续方案的不合理，而协同优化的结果有效提升线上转向接续方式的使用率，实现了再生能的高效利用。该协同优化方法能够在保证乘客服务水平和方案可实施性的前提下生成具有更低运行成本的节能优化方案，更好地实现再生能利用和车底周转连续性之间的平衡，从而实现平峰期城轨系统的高效运行和节能减排的目标。

近年来，运营商为缓解客流过饱和带来的不利影响做出了大量努力。一方面，一些运营商在基础设施、车队和人员方面进行了大量投资，以减发车的间隔，以便在高峰时段容纳更多的乘客。然而，在一些特大城市，出行需求的增长速度超过了容量供应的扩张，这进一步加剧了过度饱和的问题。事实上，在一些特大城市，高峰时段的发车间隔非常小，几乎没有空间通过增加车辆容量来缓解过度饱和。在实践中，运行效率不仅由列车时刻表决定，还受到客流装载方案（如不同 OD 对的容量分配决策^[8]）的影响。2023 年，Jinpeng Liang 等人设计了一种在线客流控制策略^[12]，对每个 OD（始发地-目的地）对的客流进行管理，使研究区间内的乘客总等待时间最小化。假设 OD 需求信息随时间顺序显示，将在线客流控制问题描述为随机动态规划（DP）。设计一种有效的在线自适应策略来指导各个阶段的实时客流控制决策。结果表明，与先到先服务（FCFS）策略相比，该方法可以显著减少乘客的预期总等待时间，并缓解地铁车站的拥堵，利用列车容量的可重复使用特性，在高峰时段运送更多的乘客。为了以最小化乘客站台等待时间为目标，Jiawei Yuan 等人建立了混合整数非线性规划模型，并设计了一种新颖的混合遗传算法^[13]。在许多大城市的轨道交通线路中，高峰时段的客流需求往往呈现过度拥挤、分布不均衡等特点。针对单条双向的城市轨道交通线路，对大小交路方案、列车时刻表和车底运用计划进行了协同优化，以最小化乘客站台等待时间为目标，分别考虑了交路选择、发车间隔、列车容量、车底衔接、车底数量等约束条件，构建了混合整数非线性规划模型。根据模型特点，设计了一种新颖的混合遗传算法，基于北京地铁 6 号线的历史数据，验证了该方法的有效性。

公交设施选址也是一项复杂的工程，王晓辉的研究^[14]以沈阳市浑南新区为背景，在考虑成本因素后，基于模糊软集理论的方法，分析了关键指标确定最佳选址。在动态流量下，同时还考虑设施的弹性，风险事件（如自然灾害、人为破坏等）会导致道路交通系统服务水平下降，良好的计划可以将期间的后续影响降到最低。关键基础设施（Critical Infrastructure, CI）的弹性对于整个社会抵御、响应风险事件并从中快速恢复至关重要。系统弹性（也称韧性）需要从多个维度、采用复合指标进行度量，而恢复力可视为评价系统韧性的维度之一。2020 年 Tingting Zhao 和 Yu Zhang 的研究^[15]重点研究了交通运输系统恢复力的评价和优化问题，并

将其建模为双层双目标优化问题。采用加权和法求解该双目标优化问题的帕累托前沿。该研究将修复计划中的优化问题建模为双层双目标优化问题，优化目标为最小化总行程时间和系统中未满足的出行需求，提出了该优化问题的有效求解算法，将该方法应用于典型路网，说明了应用该方法解决实际路网中修复计划优化问题的具体过程，同时验证了该方法的有效性，并从目标空间分析的角度进一步对该双目标优化问题的实证结果进行了阐释。在该研究中，上层优化问题的目标是通过确定需要优先修复的路段和相应的通行能力恢复等级来最小化总行程时间和系统中未满足的出行需求。下层基于弹性用户均衡问题对居民的出行行为进行建模，以实现事件发生后交通系统供给侧退化的通行能力和需求侧已经基本恢复正常的出行需求之间的供需不平衡问题的模型化表达和量化分析。现实中常需要综合运用枢纽选址、随机规划等基础理论，Haifeng Zhang 等人的研究^[16]针对面向多模式的枢纽选址和网络设计问题，分别在不确定需求和不确定运输成本情形下，建立集成优化方法对多模式货物运输系统中的枢纽与枢纽弧选址、运输模式选择以及运输路径分配进行综合研究。随机需求模型本质等价于一个确定性期望值问题，此外，该研究进一步提出了不确定运输成本下的两阶段随机规划问题，而对于随机运输成本模型，并不等价于相应的确定性期望值问题。为了有效求解随机运输成本模型，设计基于抽样平均近似技术的 Benders 分解算法。最后，利用 TR 数据集对随机规划模型和基于抽样平均近似技术的 Benders 分解算法进行检验，并通过对比分析探讨随机解的价值。结果表明，相较于随机需求模型，多模式枢纽网络拓扑结构对于随机运输成本更为敏感；与确定模型相比，随机模型能够有效抵御运输成本不确定性带来的影响，并降低枢纽网络的总成本。

1.3 研究内容

1、了解在大数据时代，机器学习以及运筹优化在交通优化中的应用的国内外研究现状，其中着重了解基于数据挖掘和机器学习的乘客出行模式分析和技术运筹优化的公交调度方法。深入学习交通模拟软件 SUMO，数值求解包 Python-MIP 的具体使用方法，以支撑后续实验研究过程。

2、深入研究 LSTM、Transformer 的基本原理和模型结构，对比机器学习算法与传统排队理论的区别，尝试使用不同算法对乘客出行模式进行预测。

3、深入研究 MIP（混合整数模型）和 DP（动态规划）问题的基本原理和解法，结合 Python-MIP 模块，研究如何对公交调度问题进行建模以及求出所得问题的数值解。

4、尝试在交通模拟软件 SUMO 中使用前文所得的结果，对不同场景下的结果进行模拟，对比不同结果之间的效率，使用可视化方法更直观、形象的展示不同方案的优劣，通过不同指标进行分析，给出优化改进方案。

1.4 论文组织结构

本文拟分为五章来对该课题进行研究，在基于数据挖掘和机器学习的基础上分析了影响乘客出行模式的因素，对交通流量预测和公交预测方面作了相关研究，具体内容如下：

第一章主要介绍研究背景、研究意义、国内外研究现状、主要研究内容和结构，并梳理了相关研究的发展历程；最后引出了本文的主要研究内容。

第二章主要介绍研究中需要使用的 LSTM、Transformer 算法，以及运筹优化中公交选址、班次调度的不同优化算法，介绍常用数值求解软件和交通模拟软件。

第三章对公交的刷卡数据、线路数据、天气数据等进行处理，因刷卡数据不包含刷卡站点、下车站点等信息通过数据挖掘和机器学习算法对乘客的行为进行挖掘，本章将介绍设计的相关算法。通过数据挖掘，分析影响乘客行为模式的特征。对各个算法的效果和特点进行分析。第四章通过前文的分析结果建立数学模型，从成本、效益、乘客舒适度等多个角度出发进行数值仿真模拟，对比不同优化结果的有效性。

第五章总结本文工作、并且展望未来研究方向。

2 研究相关理论

2.1 排队论

排队是在日常生活中经常遇到的现象，如顾客到商店购买物品、病人到医院看病常常要排队。此时要求服务的人数超过服务机构的容量，即到达的顾客不能立即得到服务，因而出现了排队现象。这种现象不仅在个人日常生活中出现，电话局的占线问题，车站、码头等交通枢纽的车船堵塞和疏导，故障机器的停机待修，水库水量的存储调节等都是有形或无形的排队现象。由于顾客到达和服务时间的随机性，可以说排队现象几乎是不可避免的。如果增添服务设备，就要增加投资或发生空闲浪费；如果服务设备太少，排队现象就会严重，对顾客个人和对社会都会带来不利影响。因此，管理人员必须考虑如何在这两者之间取得平衡，经常检查目前处理是否得当，研究今后改进对策，以期提高服务质量，降低成本^[17]。

排队模型常表示成 $X/Y/Z/A/B/C$ （相继到达的间隔时间/服务时间的分布/服务台的数目/容量限制/客源数目/服务规则）形式，对于一个公交站台可以表示成标准的 $M/M/1$ 模型，该系统的特点：

(1) 输入过程——顾客源是无限的，顾客单个到来，相互独立，一定时间的到达数服从泊松分布，到达过程已是平稳的。

(2) 排队规则——单队，且对队长没有限制，先到先服务。

(3) 服务机构——单服务台，各顾客的服务时间是相互独立的，服从相同的负指数分布。此外，还假定到达间隔时间和服务时间是相互独立的。

已知 $M/M/1$ 模型到达规律服从参数为 λ 的泊松分布，服务时间服从参数为 μ 的负指数分布，给出 t 时刻状态为 n （系统中有 n 个顾客）的概率 $P_n(t)$ ，系统的状态方程可以表示为：

$$\begin{cases} -\lambda P_0 + \mu P_1 = 0 \\ \lambda P_{n-1} + \mu P_{n+1} - (\lambda + \mu) P_n = 0 \quad n \geq 1 \end{cases} \quad (2.1)$$

式 2.1 是关于 P_n 的差分方程，它表明的各状态之间的转移关系如图 2.1 所示。

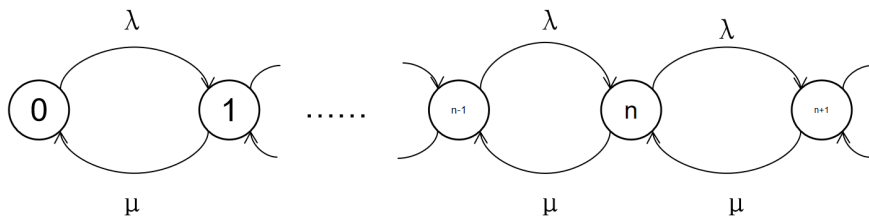


图 2.1 M/M/1 模型状态转移示意图

参考文献:

- [1] 南京市交通运输局. 2022 年一季度南京交通经济运行综合分析报告[EB/OL]. 2022[2022-06-10]. http://jtj.nanjing.gov.cn/njsjtysj/202206/t20220610_3442578.html.
- [2] 高德地图. 2022 年度中国主要城市交通分析报告[EB/OL]. 2023[2023-01-18]. <https://report.amap.com/share.do?id=a184b07685866c850185c2e9da200007>.
- [3] 百度地图. 2022 年度中国城市交通报告[EB/OL]. 2023[2023-03-01]. <https://jiaotong.baidu.com/cms/reports/traffic/2022/index.html>.
- [4] Jiang W, Ma Z, Koutsopoulos H N. Deep learning for short-term origin-destination passenger flow prediction under partial observability in urban railway systems[J]. Neural Computing and Applications, 2022, 34: 4813-4830.
- [5] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[Z]. 2017. arXiv: 1706.03762.
- [6] 高榕, 万以亮, 邵雄凯, 等. 面向改进的时空 Transformer 的交通流量预测模型[EB/OL]. (2021-03-14)[2023-04-12]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20220618.1141.008.html>.
- [7] Chen C, Liu Y, Chen L, et al. Bidirectional spatial-temporal adaptive transformer for urban traffic flow forecasting [J]. IEEE Transactions on Neural Networks and Learning Systems, 2022: 1-13.
- [8] Chierici A, Cordone R, Maja R. The demand-dependent optimization of regular train timetables[J]. Electronic Notes in Discrete Mathematics, 2004, 17: 99-104.
- [9] Nair V, Bartunov S, Gimeno F, et al. Solving mixed integer programs using neural networks[Z]. 2021. arXiv: 2012.13349.
- [10] Liu R, Li S, Yang L, et al. Energy-efficient subway train scheduling design with time-dependent demand based on an approximate dynamic programming approach[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(7): 2475-2490.
- [11] Mo P, Yang L, D' Ariano A, et al. Energy-efficient train scheduling and rolling stock circulation planning in a metro line: A linear programming approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(9): 3621-3633.
- [12] Liang J, Zang G, Liu H, et al. Reducing passenger waiting time in oversaturated metro lines with passenger flow control policy[J]. Omega, 2023, 117: 0305-0483.
- [13] Yuan J, Gao Y, Li S, et al. Integrated optimization of train timetable, rolling stock assignment and short-turning strategy for a metro line[J]. European Journal of Operational Research, 2022, 301(3): 855-874.
- [14] 王晓辉. 城市公共交通设施选址方法研究[D]. 沈阳: 东北大学, 2012.
- [15] Zhao T, Zhang Y. Transportation infrastructure restoration optimization considering mobility and accessibility in resilience measures[J]. Transportation Research Part C: Emerging Technologies, 2020, 117: 102700.
- [16] Zhang H, Yang K, Gao Y, et al. Accelerating benders decomposition for stochastic incomplete multimodal hub location problem in many-to-many transportation and distribution systems[J]. International Journal of Production Economics, 2022, 248: 108493.
- [17] 《运筹学》教材编写组. 运筹学[M]. 北京: 清华大学出版社, 2012.