本文下载日期:[218.19.145.32]时间:2020年10月20日，时间:00:08

美国运筹学与管理科学研究所(INFORMS)，位于美国马里兰州



**管理科学**

出版细节，包括作者说明和订阅信息:http://pubsonline.informs.org

数据驱动的急诊科患者调度:一种鲁棒-随机混合方法

何双池，沈melvyn，张梅林

**引用本文:**

何双池，张梅林(2019)基于数据驱动的急诊科患者调度:一种混合鲁棒-随机方法。管理科学65(9):4123 - 4140。https://doi.org/10.1287/mnsc.2018.3145

**完整的使用条款和条件:https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-条件**

本文只可用于研究、教学和/或私人研究目的。除非另有说明，否则未经出版商明确批准，禁止商业使用或系统下载(由机器人或其他自动程序)。欲了解更多信息，请联系permissions@informs.org。

发行者不保证或保证文章的准确性、完整性、适销性、适合特定目的或不侵权。本文中对产品或出版物的描述或引用，或广告的包含，既不构成也不暗示对该产品、出版物或服务的保证、背书或支持。

版权所有©2019,INFORMS

**请向下滚动查看文章——它在后面的页面上**



INFORMS拥有来自近90个国家的12,500名成员，是最大的运筹学(O.R.)和分析专业人员和学生的国际协会。INFORMS为各种类型和规模的个人专业人员和组织提供了独特的网络和学习机会，以更好地理解和使用手术室和分析工具和方法来转换战略愿景和实现更好的结果。

欲了解更多关于INFORMS的信息，请访问http://www.informs.org

IMAGE

**http://pubsonline.informs.org/journal/mnsc**

**管理科学**

**Vol. 65 No. 9, 2019年9月，pp. 4123-4140 ISSN 0025-1909 (print)， ISSN 1526-5501(在线)**

**数据驱动的急诊科患者调度:一种鲁棒-随机混合方法**

**Shuangchi他,a 梅尔文Sim卡,b 梅林天虹张c**

**a** 新加坡国立大学工业系统工程与管理系，新加坡117576; **b** 新加坡国立大学新加坡国立大学商学院分析与运营系，新加坡119245; **c** 新加坡社会科学大学商学院，新加坡599494



**联系人:heshuangchi@nus.edu.sg, http://orcid.org/0000-0003-4107-3946 (SH);**melvynsim@nus.edu.sg,

http://orcid.org/0000 - 0001 - 9798 - 2482 (MS);zhangmeilin@suss.edu.sg, http://orcid.org/0000 - 0003 - 2880 - 8223 (MZ)

**收稿日期:2015年11月22日修回日期:2017年2月19日;**录用日期:2018年5月24日

**提前在网上发表:2019年5月1日**

**https://doi.org/10.1287/mnsc.2018.3145版权所有©2019 INFORMS**

**摘要**紧急护理需要充分和及时的治疗，但不幸的是，由于许多急诊科(EDs)人满为患，这一点受到了影响。为了解决这一问题，我们研究了急诊科的患者排班问题，以便在最大的可能性下集体满足每个患者的上门服务时间和住院时间的强制性目标。利用急诊科的患者流量数据，我们提出了一种混合鲁棒-随机方法来制定患者调度问题，该方法考虑了实际的特征，如时变的患者到达过程、一般会诊时间分布和多异质性医生。与传统的最大化目标达成联合概率的计算方法相比，该混合方法提供了一个计算友好的公式，可以得到令人满意的病人调度问题的解。这个公式使我们能够开发一个动态调度算法，为每个可用的医生为下一个病人提供建议。在数值实验中，所提出的混合方法优于样本平均近似方法和渐近最优调度策略。

**历史:被Yinyu Ye接受，优化。**

**新加坡教育部学术研究基金[Grant MOE2017-T2-1-012]资助。**M. Sim的工作得到了新加坡教育部社会科学研究部的部分支持[MOE2016-SSRTG-059]。

**补充材料:e-companion可在https://doi.org/10.1287/mnsc.2018.3145获得。**

**关键词:医疗运营•患者调度•鲁棒优化•随机规划•混合整数规划•排队网络**

**1.介绍**

急诊科(ED)拥挤和随之而来的延误已经成为一个世界性的问题，并受到政府、公共媒体和学术界的相当大的关注。急诊科的拥挤降低了急诊护理的质量和获得机会，使病人面临巨大的治疗错误的风险。大量研究表明，急诊科人群拥挤与发病率和死亡率增加之间存在关联(McHugh, 2013年)。对医院来说，急诊科拥挤损害了他们的公共声誉，并导致收入损失，因为救护车转移，病人离开而不被发现。正如Rabin et al.(2012)所指出的那样，大范围的拥挤也阻碍了医院实现国家安全和质量目标的能力，损害了医疗体系，并限制了地区应对灾害的能力。在许多国家，政府机构密切监测医院的急救工作，并定期公布一些关键指标。例如,

美国医疗保险和医疗补助服务中心(CMS)在其“医院比较”网站上公布了4000多家医院及时紧急护理的质量指标;其中一些措施包括在CMS的绩效工资计划中。为了解决拥挤问题，政府和管理机构可以为紧急护理设定强制性目标。2005年，英格兰国家卫生服务机构强制要求98%的急诊科病人必须在到达医院四小时内接受治疗，或者出院回家，或者住院。这一“四小时规则”的实施大大提高了在急诊科花费少于四小时的患者的比例，从2002-2003年的77.3%提高到2008-2009年的97.2% (Weber et al. 2011)。

Hoot和Aronsky(2008)总结了急诊科拥挤的常见原因，包括急诊护理需求的增加、医院床位容量不足、操作效率低下等。有效的患者流量管理有望成为解决方案

4123

**He, Sim, and Zhang:数据驱动的急诊科患者调度**

管理科学，2019,vol. 65, no. 4。9, pp. 4123-4140，©2019 INFORMS

4124

过度的病人延误，而没有直接扩大容量。为了评价紧急护理的及时性和效率，国家质量论坛认可了住院时间、上门服务时间(即例如，病人在接受医疗保健服务之前在急诊科度过的时间)，以及离开时不被视为质量指标(Welch et al. 2011)。因为不被看到就离开的百分比与患者上门治疗的时间密切相关，我们将这两个时间指标视为主要的性能问题。一般来说，根据每个患者的临床紧急情况，上门服务的时间应保持在一定的安全限度以下。例如，广泛使用的急诊严重程度指数(Emergency Severity Index)将急诊科病人根据他们的急性程度和所需的医疗资源分为五组;推荐的上门到提供者的时间目标范围从复苏患者的“立即”到不太紧急的患者的“一到两个小时”(Gilboy等人，2011年)。临床和操作需求都对患者流量管理施加了严格的时间限制。

本文的研究重点是急诊科的病人排班问题。从建模的角度来看，ED可以被看作一个排队网络，以医疗单元为节点;病人是顾客;病床、医务人员和设备是服务器(Armony等人，2015年)。除了优先考虑的客户和对时间敏感的服务要求外，该网络的特点是客户经常返回(即:在与医生进行初步咨询后，大多数患者会进行医学检查，并在最终出院或住院之前回到同一位医生那里)。尽管急诊医生被要求为广泛的疾病和伤害提供治疗，但他们的专业知识和工作效率各不相同。换句话说，这个网络的服务器是异构的。当有多个病人等待就诊时，他们各自的医生和会诊顺序必须仔细安排，以满足严格的上门服务和住院时间目标。但由于网络结构复杂、服务器异构、患者到达过程高度不确定、服务需求时间敏感等特点，对患者调度问题的解决提出了挑战。

为了在实际应用中解决这个问题，我们提出了一种混合鲁棒-随机方法来利用患者流量数据进行实时患者调度。我们的目的是最大限度地提高上门治疗时间和住院时间在强制目标范围内的患者比例。由于患者到达模式是高度可变的，我们避免对未来的患者到达做出假设，如到达率和到达间隔时间分布。利用现有患者的数据，动态调度算法将确定下一个患者

只要有医生，就去看他。为了及时提出建议，调度算法必须足够高效。

我们可以通过最大化所有等待患者达到延迟目标的联合概率来构造一个优化问题来获得时间表(见式(6)和式(7)及相关讨论)。以目标达成的联合概率为目标，Charnes和Cooper(1963)首先研究了这一优化问题，并将其称为P模型。然而，P模型公式在实践中并没有得到广泛的应用，部分原因是评估联合概率需要在高维上进行积分，这通常是难以计算的，更不用说解决相关的非凸优化问题了。

为了解决这个问题，我们通过考虑一组不确定性集，将鲁棒优化的特征纳入我们的公式。与给定的时间表相关联，家庭中设定的每个不确定因素包括患者在不违反强制延迟目标的情况下所能接受的可行会诊时间。不像传统的鲁棒优化公式，其中不确定集是固定的，混合方法在家族中搜索具有最大可能性的所有咨询时间是可行的不确定集。所得到的不确定集所对应的调度是混合公式的最优解。由于计算上的原因，我们将不确定集家族限制为超矩形的集合。然后，在会诊时间独立的假设下，所有等待患者达到延迟目标的联合概率简单为每个患者达到其自身延迟目标的边际概率的乘积。在这种情况下，联合概率的计算不涉及高维积分，可以大大提高调度算法的计算效率。在数值实验中，鲁棒-随机混合方法优于样本平均近似(SAA)方法和渐近最优策略;详情见第7节。

鲁棒-随机混合方法具有重要的实用和方法论意义。首先，虽然混合公式本质上是一个混合整数规划，解决这个问题实际上是有效的，并允许在EDs的实时调度。作为一种由数据驱动的动态方法，它支持实际功能，例如时变的病人到达过程、一般会诊时间分布和异质性的医生。在关于排队网络调度的文献中，这些特性在现有的网络模型中通常是缺失的。可能导致现有调度策略在实际应用中效果不佳。第二，混合公式代表了解决P模型问题的另一种视角，

**何sim, Zhang:基于数据驱动的急诊科患者调度研究，2019,vol. 65, no. 4。**9, pp. 4123-4140，©2019 INFORMS

4125

其目标是最大化随机摄动线性约束的可行性概率。可以想象，混合公式可以以低得多的计算费用产生接近最优解。如7.2节中的数值例子所示，我们的方法可以提供SAA方法的一种高效的替代方法。除了病人的时间安排，类似的问题也出现在其他具有时间敏感性服务要求的随机系统中;第8节有更多的讨论。

本文的其余部分组织如下。相关文献在第2节中进行了回顾。我们在第3节介绍了EDs的排队网络模型。在第4节中，我们提出了一种基于混合鲁棒-随机公式的病人调度问题的可处理方法。这个混合公式在第5节中被转换成一个混合整数规划。在第6节中，通过引入额外的延迟约束，将混合公式引入动态调度框架中，使我们能够按照随机的病人到达过程顺序地解决病人调度问题。我们在第7节中提供了一个全面的基于数据的数值研究，其中混合方法与SAA方法和渐近最优调度策略进行了比较。本文在第8节进行了总结，并讨论了一些潜在的应用和未来的研究。我们将非预期安排的构造、所有证明和额外的模拟结果留给e同伴。

让我们用常用的符号来结束这一节。标量和向量分别用小写字母和粗体字母表示。书法字母用于集合，例如(，我们使用|(|表示集合的基数。随机变量和向量用波浪符号表示，如s和s。我们假设所有的随机变量和向量都定义在一个公共概率空间中，其中P(a)是事件a发生的概率。我们保留E(s)作为随机变量s ~的期望。

**2.相关文献**

我们概述相关研究，以便在文献中定位我们的工作。关于病人流量管理和排队网络优化的文献都是广泛和良好的建立。我们不打算详尽无遗。

出于分析和控制的目的，EDs通常被建模为排队网络。虽然大多数研究都是基于模拟的(见Connelly和Bair 2004, Sinreich和Marmor 2005，以及其中的参考文献)，但在分析研究中使用了几种简化的排队模型。例如，为了确定医生的人员配备水平，Green等人(2006)用时变Erlang-C模型描述了ED占用过程，de Bruin等人(2010)用时变Erlang-B模型描述了ED占用过程。

Yom-Tov和Mandelbaum(2014)利用患者可能多次返回同一位医生的特点，提出了一种改进的占用过程Erlang-R模型。Saghafian等人(2012)分析了患者分流的实践(即:他还提出了一种改进的分流方案。Saghafian等(2014)提出了一种基于临床急迫性和治疗复杂性的新型分诊系统，以提高患者的安全性和操作效率。

急诊科的患者排班由Huang et al.(2015)研究，在文献中与我们的工作最相关。在他们的论文中，ED被建模为一个具有服务截止时间和反馈路径的多类排队网络。作者提出了一种简单而高效的调度策略，能够在维持可接受的上门服务时间和缓解拥堵之间取得平衡。通过大流量分析，他们证明了在一个简化的设置下，他们所提出的调度策略是渐近最优的，以减少总拥塞成本，且受门到供应商时间的约束。这个调度策略是我们混合方法的重要基准;第7.3节对这两种方法进行了比较。

尽管前面提到的排队模型能够表示ED的基本操作特征，但它们可能过于简单，无法捕捉一些显著特征。对于一个可分析处理的排队模型，可能需要概率假设，如指数到达间隔和服务时间分布，平稳到达过程和同构服务器。Bertsimas et al.(2011)指出，如果没有这些假设，排队网络的性能分析在很大程度上是无法解决的。然而，由于环境复杂且变化频繁，这些假设可能并不合适。可以想象，在这些假设下获得的控制政策在实践中不一定能很好地工作。相反，提出的混合公式不依赖于这些假设。因此，我们希望这种数据驱动的方法能够更好地适应ED环境。

即使排队网络模型是解析可处理的，仍然很难找到具有延迟或吞吐量时间约束的最优动态控制策略。文献中的大多数研究关注的是在某种渐近意义上可以证明最优的简单策略(例如，Doytchinov等人2001年，Plambeck等人2001年，Maglaras和Van Mieghem 2005年，以及Huang等人2015年)。前面提到的一些简单假设，以及繁忙的交通条件，对于保持渐近最优性是必要的。当简单化的假设不满足时，这些政策的性能可能不是最优的。此外，这些策略的控制动作依赖于服务时间

**He, Sim, and Zhang:数据驱动的急诊科患者调度**

管理科学，2019,vol. 65, no. 4。9, pp. 4123-4140，©2019 INFORMS

4126

分布只通过它们的一阶矩。为了使这些策略接近最优，延迟或吞吐量时间的最后期限必须高于服务时间的数量级，这在ED设置中可能不是一个合理的假设。相反，混合鲁棒-随机方法允许多个异构服务器、时变到达过程和任意交通条件。从患者流量数据中获得的分布信息被充分地用于构建不确定性集。换句话说，混合方法可以利用整个服务时间分布，这比以前的调度策略具有相当大的优势。在7.3节的数值实验中，即使ED在交通拥挤的情况下，混合方法也优于Huang et al.(2015)提出的渐近最优调度策略。

在鲁棒优化框架下，Bertsimas等人(2011)、Bandi and Bertsimas(2012)和Bandi等人(2015)研究了排队网络的性能分析。在他们的论文中，到达和服务时间的随机性是用概率论中的极限定律用多面体不确定性集建模的。具体来说，Bertsimas et al.(2011)考虑了迭代对数定律，Bandi和Bertsimas(2012)和Bandi et al.(2015)考虑了构造不确定性集的广义中心极限定理。利用这种鲁棒优化方法，得到了排队网络的性能边界。尽管我们的方法也受到了稳健优化的启发，但它来自一个完全不同的视角。与传统的鲁棒优化公式不同的是，不确定性集被指定为固定的约束条件，我们的方法研究了一系列的不确定性集，并寻找在这些不确定性集中“最大化”不确定性集的计划。从这个意义上说，得到的调度方案是最“稳健”的病人调度方案。

我们的混合鲁棒-随机方法与Zhang等人(2017)的并行独立研究有一些相同的特点，他们认为具有可调不确定性集的约束线性系统的鲁棒最优控制。他们的问题的形成是由电网的备用金驱动的，在电网中，为了节约成本，需要定期调整电力的备用金容量，而不牺牲必要的电力消耗。在它们的公式中，使用一系列可调的不确定性集来表示备用容量，从而成为决策变量，如我们的方法。Zhang等人(2017)也研究了特殊几何形式的不确定性集，以使其优化问题易于处理。尽管有这些相似的特征，但我们的研究与他们的工作在以下几个方面有很大的不同。从建模的角度来看，可调不确定性集出现了

而在我们的方法中，超矩形集主要作为解决P模型问题中可行集的启发式代理。与联合概率测度相关联，这些超矩形集不应该被理解为通常意义上的不确定性集。从方法论的角度，他们的论文的主要主题是如何限制不确定性集和允许的政策，以仿射结构，从而产生的公式成为一个凸优化问题。相反，我们将病人调度问题转化为一个可处理的混合整数规划，依据的事实是，在每个超矩形内，最坏的情况只发生在单个边界点(在第4节定理1中讨论)。值得一提的是，作为决策变量，两种研究中的不确定性集合都受控制策略的联合约束。换句话说，可调不确定性集合的家族可能会随着具体的政策而改变。这类似于Spacey等人(2012)提出的稳健优化公式，他们研究了具有多实例化的软件划分(即。例如，将计算机程序的代码段分配给多个执行位置，以使整个程序运行时间最小化)。在他们的问题中，唯一的决策变量是软件分区，它也决定了位置感知控制流的不确定性集合。与我们的研究不同，它们的不确定性集是不可调的，因此，它不是优化问题的决策变量。

**3.控制排队**

**网络模型**

我们使用排队网络来模拟急诊科，它是由一个集中的病人调度系统控制，以满足门到提供者的时间和住院时间的要求。根据当前患者的状态，排班系统将为每个医生对下一个患者进行顺序的建议。

急诊病人的一般流程是按照以下流程进行的。病人以随机和非平稳的方式到达急诊科。注册后，他们将由一名护士进行分类，并根据他们的视力水平和其他问题分配到几个紧急小组。每个急症组患者的上门服务时间应保持在规定的安全限度以下，而急症组的安全限度可能各不相同。然后，病人会呆在等候区，直到医生叫他们去看医生。这些病人将被称为新病人。在最初的会诊后，一些病人可能会离开急诊科，而另一些病人可能会接受诊断测试，如x光和血液测试，或接受护士的治疗。当检测结果准备好或治疗完成时，病人

**何sim, Zhang:基于数据驱动的急诊科患者调度研究，2019,vol. 65, no. 4。**9, pp. 4123-4140，©2019 INFORMS

4127

将返回等候区，成为返回病人，等待同一位医生的检查。在最终出院或住院之前，病人可能会多次见同一位医生。

这个排班系统将决定将病人分配给每个医生以及他们会诊的顺序。新病人可以分配给任何可用的医生，而返回的病人必须由他最初咨询的医生看。我们假设调度系统不管理等待护士检查或治疗的患者，因为这些患者通常是在先到先得(FCFS)的基础上得到服务的。病人排班系统对于缓解急诊科拥挤是至关重要的，因为简单的优先顺序规则无法平衡上门服务的时间和住院时间。如果医生优先考虑新患者，以减少上门就诊的时间，返回的患者就不得不花更多的时间等待，并排起长队。根据利特尔定律，平均停留时间会延长。特别是，需要多次咨询的患者将有很长的总等待时间，这可能会在住院时间的分布上产生一个长尾。当急诊科人满为患时，这些病人的住院时间很可能会超过规定的目标。优先考虑返回患者可以有效缩短候诊区的排队时间，从而缩短住院时间。然而，这种策略将不可避免地延长新患者从诊所到医院的时间，使他们面临治疗延误的风险。此外，为了保持操作效率，在决定下一个病人就诊时，必须考虑每个医生的专业知识。一般来说，在门诊到医生的时间和住院时间都受到限制的情况下，很难找到一个合理安排病人时间的经验法则。

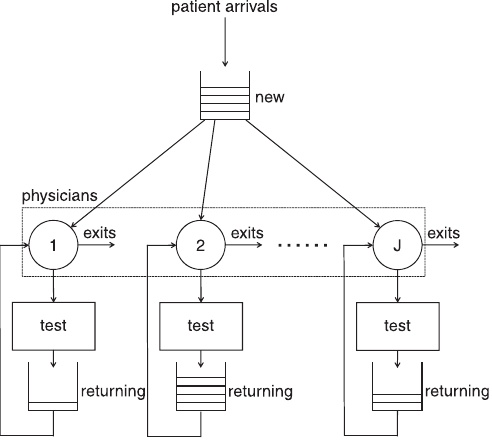
受控排队网络模型如图1所示。当一个医生完成一次会诊时，排班系统确定要见的下一个病人，当一个新病人来到候诊区寻找至少一名免费医生时，排班系统确定他的医生。让我们在这样的时刻，考虑一下此时此刻的ED。。设)是医生的集合，1是新患者的集合，#是被观察到的患者的集合，5是返回的患者的集合，为了便于表示，这里抑制了对t的依赖。对于j∈)，我们使用#j 表示医生j和医生5所诊治的一组病人j 表示j医生要看的一组返回病人。然后，

#吗?⋃j∈)＃j 和5 ?⋃ j∈)5j．此外,#j ？Ø如果



只有一个病人。让0 ?1∪5是候诊区的一组患者(?可以是急诊室里的病人，不包括那些送去检查或治疗的。

**图1所示。**急诊科病人流的排队网络模型



对于i∈(和j∈)，设s属于ij 如果i∈#，则为病人i的会诊时间js˜ij 被解释为病人i的剩余会诊时间，因为他正在由医生j诊治。我们假设{sij : i∈(，j∈)}是一组相互独立的随机变量，使用Fij表示s ~ ij的累积分布函数。因为即使i∈(是固定的，Fij 对于不同的j∈可能不同)。每个Fij 可以通过医生j的会诊时间记录来估计，这可能取决于医生的专业知识以及患者的状态(新或复诊)、分诊信息、初步诊断等。在我们的实施中，Fij被认为是一个被选择的咨询时间样本的经验分布函数。因此，我们假设每个s属于ij是一个离散的随机变量，其值取自一个有限的正数集合

6ij ？{年代ij(1)，…，sij(Nij)}。我们使用的年代ij 和s¯ij 分别表示6ij中的最小和最大的数。



让年代˜?(s ~ ij)i∈(，j∈)是所有这些协商时间的随机向量。那么，s ~从产品空间6 ?∏我∈(∈)6 ij。

将等待的患者分配给医生是由函数φ: 0→)指定的，其中ϕ(i)是患者i的医生。因为返回的患者必须由其最初的医生查看，所以该分配应该满足要求



(1）

排序决定由对应的Φ: 0→3(0)指定，其中3(0)为0的幂集，Φ(i)为在患者i之前由同一位医生诊治的患者集合。

ϕ(k) ?其中n (i) for k∈Φ(i) and i∈0。（2）

**He, Sim, and Zhang:数据驱动的急诊科患者调度**

管理科学，2019,vol. 65, no. 4。9, pp. 4123-4140，©2019 INFORMS

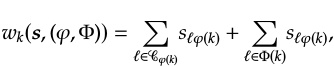
4128

对于由同一名医生看病的患者，相关的Φ(i)形成一个嵌套集集合:对于i, k∈0，这样的φ (i) ?其中n (k)，我们必须



（3）

如果这对分配和排序决策(ϕ， Φ)满足(1)-(3)，则称为可接受的调度。我们使用!表示所有容许的时间表的集合。对于给定的调度(ϕ， Φ)，∈!和给定的协商时间s ?(sij)i∈(，j∈)∈6，则病人k∈0的等待时间可由



（4）

其中右边的第一个和是医生为患者看病的剩余会诊时间(如果# φ (k)，这是零?Ø)，第二和为等待患者在患者k之前的总会诊时间。

我们假设每个病人i∈0都有一个延迟目标τi．对于新病人i∈1，τi 从t到他的等待时间超过他的门到供应商时间的安全限制的时间。对于返回的病人i∈5，τi 由计划系统指定他的停留时间以达到强制目标。我们将在第6节讨论如何确定返回患者的延迟目标。排期系统需要为现有患者找到一个可接受的排期，在该排期下患者的等待时间不应超过延迟目标。但是，由于会诊时间是随机的，我们可能无法完全确定地做到这一点。相反，我们寻求使所有患者等待时间都在延迟目标内的联合概率最大化。

安排是一个函数π: 6→!，它将协商时间的实现映射到一个可接受的时间表。我们用9来表示所有排列的集合。在给定的安排下，我们可以用(4)来估计所有等待时间在目标范围内的联合概率。因此，通过求解以下P模型问题，可以得到一个最优的安排

马克斯·P (wi(s ~， π(s))≤τi π∈9。

（5）

然而，这个公式不能实现，因为要确定允许的时间表，需要提前知道s ~的实现。为了解决这个问题，我们应该将可行的解决方案限制在(5)非预期安排的集合内，这些安排不依赖于未来的信息来确定何时可以看病的病人。要指定非预期的安排，我们需要以顺序的方式确定分配和排序决策。设w(1)≤w(2)≤…为时

医生可以开始会诊。为了决定w(k)时间的下一个患者就诊，我们可以利用w(k)之前的患者和医生信息，包括每个医生的会诊历史、等待患者的身份、医生与当前患者相处的时间等。使用累积信息，我们可以通过递归过程定义非预期安排。因为非预期安排的构造通常是复杂的，我们把细节留给e同伴。让91 是所有非预期安排的集合。通过求解该问题，我们可以得到一个最优的非预期安排

马克斯·P (wi(s ~， π(s ~))≤τi :我∈0)

(6）



P模型问题(6)被证明是棘手的，部分原因是由于非预期安排的递归结构引起的维数诅咒。为了简化计算，我们可以进一步将可行解限制在(5)的静态排列集合内，由

FORMULA

下一个命题指出静态安排是非预期的。

**命题1。**设9、90和91分别是所有安排、静态安排和非预期安排的集合。然后,90 ⊂91 ⊂9。

由于静态安排对于协商时间的所有实现都是不变的，因此找到一个最优的静态安排将比求解(6)简单得多。等价于通过求解以下静态P模型问题得到一个最优的容许时间表:

马克斯·P (wi(s ~，µ)≤τi : I∈0)s.tµ∈!。

(7）

最优解决方案µ†?(ϕ， Φ†)到(7)指定所有等待患者的分配和排序决定。特别地，如果任意i∈0，则Φ†(i) ?Ø和#ϕ†(我)?Ø hold，患者我将是内科医生φ看到的下一个患者，应该立即发送给医生。当医生完成会诊或新病人到达候诊区，找到至少一名免费医生时，就会重复这个过程。每次，调度系统都会为可用的医生确定要看的下一个病人。

尽管求解(7)比求解(6)简单得多，但找到最优的容许调度仍然是一个相当大的挑战。在容许调度下，计算(7)的联合概率涉及

**何sim, Zhang:基于数据驱动的急诊科患者调度研究，2019,vol. 65, no. 4。**9, pp. 4123-4140，©2019 INFORMS

4129

许多变量的多维积分，这是计算上禁止的。Nemirovski和Shapiro(2006)指出，计算独立随机变量和的分布是NP-hard问题。因此，即使是找到病人等待时间的分布在计算上也会很困难。当等待患者数量较大时，我们无法在动态患者调度所需的合理时间内得到(7)的最优允许调度。因此，我们希望将重点放在一种计算上的友好方法，以获得一个接近最优的解决方案。

**4.混合动力强劲,**

**随机的方法**

考虑(4)给出的函数wk，将其定义域扩展到R+|(||)| × !。∈!，集合

FORMULA

是|(|·|)|维上的凸多面体。然后，我们可以将(7)改写为

（8）

马克斯酸处理

(P (s˜∈)-µ)

µ∈!

其最优解是使所有协商时间在相关凸多面体内的联合概率最大化的允许日程。由于很难计算P(s∈])对于一个高维的一般多面体]，因此要找到(8)的最优解将是非常麻烦的计算。然而，如果]碰巧是超矩形(例如:]?∏i∈(，j∈)[0,dij]对于某些dij≥0)，则上述联合概率可由以下公式计算得出

FORMULA

因为s ~的项是相互独立的。在这种情况下，求联合概率不涉及繁琐的高维积分。

上述观察促使我们考虑另一种方案。注意，通过(8)，我们打算找到具有s ~诱导的最大概率测度的相关凸多面体的可容许调度。如果凸多面体的概率度量是大的，我们可以期望多面体包含一个具有同样大的概率度量的超矩形子集。相反地，如果我们能找到一个包含一个“大”超矩形的凸多面体的允许时间表，我们也可以期望多面体本身是大的。因此，我们将寻找具有最大超矩形子集的凸多面体的可容许调度，而不是搜索具有“最大”凸多面体的可容许调度。因为计算概率要容易得多

在超矩形测度下，该公式可以使病人调度问题在计算上更加可亲。

我们将修改P模型问题(6)以获得一个“接近最优”的非预期安排。为此，让我们考虑由

？？

FORMULA

我们的目标是在不超过延迟目标的情况下，最大限度地提高会诊时间在超矩形集合4∈\*内的联合概率:

(s)∈4k(s， π(s))≤τk，



（9)

FORMULA

从稳健优化的角度来看，\*可以被看作是s ~的一个不确定集族，具体的不确定集4可以在\*中使用不同的安排进行调整。式(9)中的目标函数包含随机变量，而约束则基于稳健优化公式，该公式具有可调的不确定性集。因此，我们将此公式称为混合稳健-随机方法。

由于非预期安排的表示通常是复杂的，人们可能想知道混合公式是否更易于计算处理。为了解决这个问题，让我们考虑公式(9)的一般版本，它不局限于排队网络的调度问题。让σ˜吗?(˜σ1， . .， σ属于 M)为m维的相互独立的随机变量向量。k ?1，…，M，我们用Gk 表示σ ~的累积分布函数 k．我们用^来表示R的闭超矩形子集的集合M 上面有界

(即。, ^ ?{∏Mk?1(−∞,bk ]: (b1，…，bM)∈RM})。让美元



RM 美元。设θ是乘积空间R中的一个函数M × $到RL，这样，σ ?(σ1，…，σM)∈RM ν∈$，θ(σ， ν)在每个σ中都不是递减的k．那么，对于给定的η∈RL，让我们考虑以下问题:

？？

σ∈@

max P σ ~∈@ s.t θ(σ， ψ(σ))≤η，

（10）

FORMULA

定理1给出了(10)的简化形式。

**定理1。**令(b⋆，ν⋆)∈RM × $是下列问题的最优解:

M ?

ln Gk(bk）



马克斯酸处理

（11）

θ(b, k?1ν)≤η， b∈RM， ν∈$。

**He, Sim, and Zhang:数据驱动的急诊科患者调度**

管理科学，2019,vol. 65, no. 4。9, pp. 4123-4140，©2019 INFORMS

4130

设@⋆为^的超矩形子集，边值为b， ψ⋆为φ⋆(σ) ?ν⋆对于σ∈RM．(@⋆，ψ⋆)是(10)的最优解。

这个定理允许我们得到(9)的简化形式，其中静态排列是最优的。

**推论1。**设(d⋆，µ⋆)∈6 × !成为下列问题的最优解:

?？

ln作用力(dij)

马克斯

（12）



我∈(∈)

酸处理

wk(d，µ)≤τk，

FORMULA

设4⋆为\*中的超矩形不确定度集，边值为d⋆，π⋆为π⋆(s) ?µ-⋆for s∈6。(4⋆，π⋆)是(9)的最优解。

由于超矩形不确定集的存在，混合优化问题(9)具有(12)所给出的计算亲和力形式。利用这些超矩形集，计算(9)中的联合概率简化为(12)中的双重求和，而不需要高维积分。对于给定的不确定性集合，因为最坏的情况只发生在所有协商时间都取其最大值时，我们将只要求在这一单一情况下审查可接受的时间表。同样值得注意的是，现有的具有可调节不确定性集的稳健优化公式，如Bertsimas和Sim(2004)的不确定性预算和Ben-Tal等人(2004)的椭球不确定性集，可能不会像我们的混合方法那样导致更易于处理的形式。

一个可行的解决方案(4，π)到(9)可能导致不同的协商时间实现的不同的可接受时间表。从这个意义上说，它是一种可调稳健公式，类似于Ben-Tal等人(2004)提出的不确定线性规划的公式。一般来说，可调稳健公式比不可调稳健公式保守性更小，得到更好的目标值。然而，推论1暗示我们可以求解一个不可调整的公式来获得(9)的最优解，这是一个对于所有协商时间实现的可接受的时间表不变式。这是因为两种公式的最坏情况是相同的。混合优化问题(9)也可以有可调整的解，这些解是非静态的、非预期的安排。正如de Ruiter et al.(2016)所指出的，即使在最坏的情况下两者都是最优的，可调整的解决方案在平均客观值方面可能优于不可调整的解决方案。因此，可以有针对(9)的可调整解决方案，它们在患者调度方面的性能优于针对(12)的静态解决方案。因为

非预期安排的表示是复杂的，找到这种可调节的解决方案通常是困难的(e-companion中的EC.1节)。

当急诊科人满为患时，在任何可容许的时间安排下，都可能发生至少有一名病人的等待时间超过延迟目标的情况。在这种情况下，混合公式(9)没有可行解。由于(4)给出的等待时间随着每次咨询时间的增加而增加，我们可以在所有的咨询时间都占用其可接受的时间表时，通过审查可接受的时间表来确定(9)的可行性

最小的可能值(例如:,年代?(sij)我∈(∈))。更多的spe -



在Cifically，我们可以解决以下优化问题:

最小值α

钢桁架wk(s，µ)≤τk + α， k∈0,s ?(年代ij）i∈(,j∈) （13）

µ∈!

我们通过下面的命题来确定(9)的可行性。

**命题2。**设α⋆为(13)给出的α的最小值。然后，当且仅当α⋆≤0时，混合优化问题(9)存在可行解。

**5.一个混合整数程序**

等价形式(12)可以转化为一个混合整数程序，使我们可以利用现有算法解决病人调度问题。

对于i∈(，j∈)，ℓ?1，，，， |(|，让xijℓ 表示指定的医生和患者i的会诊顺序的二元变量:



如果病人I是

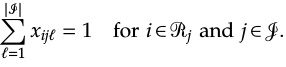
ℓj医生要看的病人，否则0。

我们为每位医生预留|(|)位置，让所有患者都可以由任何一位医生自由安排。这些二元变量必须共同满足以下约束条件。因为每个队列的第一个位置是给正在或即将被医生诊治的病人的，我们有



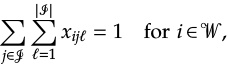
(14）

此外，因为返回的病人必须由他们最初的医生看，我们有



(15）

在允许的时间表下，每个等待病人只能被分配到一个职位:

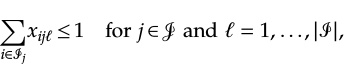


（16）

**何sim, Zhang:基于数据驱动的急诊科患者调度研究，2019,vol. 65, no. 4。**9, pp. 4123-4140，©2019 INFORMS

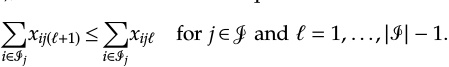
4131

每一种体位最多可容纳一名患者:



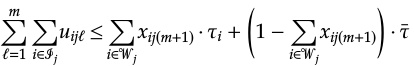
(17）

(在哪里j ？1∪#j ∪5j 是符合j医生就诊条件的患者集合。从第一个位置开始，我们必须将患者分配到每个医生的连续位置，因此，只有队列的最后才会出现空的位置。因为医生j的第ℓ个位置有一个病人，当且仅当等式在(17)中成立，上面的约束等价于



（18)

的形式。通过引入一组变量{uijℓ ≥0:I∈(j， j∈)，ℓ?1， . . .， |(|−1})，我们可以把(20)写成两个独立的不等式:



对于j∈)和m ?1， . .， |(|−1 (21))

FORMULA

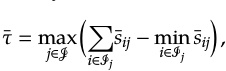
相当于



(22）

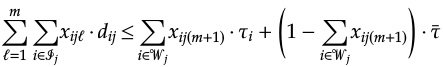
可以检验(14)-(18)等于(1)-(3)。换句话说，每个允许的时间表可以由一组二元变量{xijℓ : I∈(，j∈)，ℓ?1，…，|(|}满足(14)-(18)。

考虑(12)中的延迟约束。把



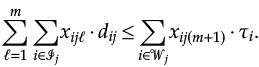
(19）

在年代¯ij s ~的最大值是ij 可以采取。那么，τ¯是病人等待时间的上界。给定一组满足(14)-(18)的二进制变量，可以将(12)中的延迟约束写为

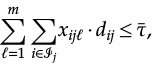


对于j∈)和m ?1， . .， |(|−1，(20))

在0j ？1∪5j 组病人等待资格是被医生j。在这种不平等,左边是医生j的时间需要完成患者的m位置或等待时间,直到医生开始服务(m + 1)圣的地位。如果病人处于(m+1) st位，i∈0jxij(m+1) ?1，不等式(20)变成



因为右边的和等于延迟目标，这个不等式是(m + 1)st病人的延迟约束。如果(m + 1)st位没有病人，i∈0jxij (m + 1) ?(20)右边的第一个和变成了0。不等式(20)的结果是



它总是根据τ¯的定义成立的。为了计算方便，让我们规范地表示(20)

然后，不等式(21)和(22)以规范形式指定延迟约束。

我们得到了(14)-(18)，(21)，(22)所给出的患者调度问题的约束，它们等价于(12)中的约束。根据下面的定理，我们可以将混合优化问题转化为一个混合整数规划。

**定理2。**让克ij(n) ?ln Fij(年代ij(n))对于I∈(，j∈)且

FORMULA

写成如下混合整数程序:

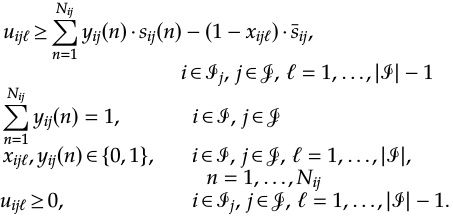


马克斯

FORMULA

限制(14)-(18)及(21)

酸处理



(23）

**6.动态调度**

**混合方法**

病人到达急诊科形成了一个随机过程。为了使调度系统做出相应的顺序决策，所提出的混合方法必须被纳入一个动态调度框架。我们假设，当一个医生完成一次会诊，或者一个新病人来到等候区寻找至少一名免费医生时，就会触发决策迭代。每次，调度系统都会为现有的医生推荐下一个病人就诊。

为了利用混合方法解决动态调度问题，我们需要在触发决策迭代时确定等待患者的延迟目标。

**He, Sim, and Zhang:数据驱动的急诊科患者调度**

管理科学，2019,vol. 65, no. 4。9, pp. 4123-4140，©2019 INFORMS

4132

新病人最关心的是他们上门看病的时间，而归国病人最关心的是他们住院的时间。设在t时刻触发决策迭代。对于i∈0，设tiDi和Ki 分别是病人i的到达时间、上门服务时间的安全限制和住院时间的强制目标。对于新病人i∈1，我们取延迟目标τi ？Di −(t−ti)，也就是病人从诊所到医生的时间超过了安全限制。假设一个病人最多可以返回同一位医生处B次。为了确定归国患者的延迟目标，我们选择B正数(Ti1…，tiB)满足

FORMULA

从病人i到达等候区到医生第(m + 1)次见他为止的强制时间限制。如果返回的病人i∈5在(m + 1)st时间内等待被看见，则取延迟目标τi ？Tim −(t−ti）.施加额外的延迟约束使我们可以通过顺序求解(9)进行动态调度。在额外的延迟要求下，还可以将患者个次就诊的等待时间返回维持在一个合理的水平，这可能会进一步提高患者的安全性和满意度。这些附加强制限制的特定值将影响动态调度的性能。如果能够根据急诊科的充血情况在每次迭代中调整返回患者的延迟目标，那将是非常理想的。因为为每个返回的病人找到最优的延迟目标通常是困难的，我们将在我们的实现中使用启发式方法来确定这些参数;第7.3节有更多的细节。

给定当前等待患者的延迟目标，调度系统首先通过求解(13)来确定混合优化问题(9)的可行性(13也可以按照第5节的步骤转化为混合整数规划)。如果(9)的可行集非空，调度系统将求解混合整数方案(23)，该方案的最优解指定了可用医生要治疗的下一个病人。如果问题(9)是不可行的，则在我们的实施中将使用求解(13)得到的可接受的时间表。在这种情况下，假设每个等待患者的会诊时间取最小值。所获得的容许时间表是所有等待病人中等待时间最长的时间最小的时间表。病人就诊的建议是根据这一可接受的时间表作出的。

我们想指出的是，这种迭代调度程序在本质上是短视的。因此，从连续迭代中得到的容许时间表可能是不一致的(即。，连续分配决策可能不满足Bellman的最优原则)。由于动态规划的公式将是棘手的，时间不一致是实际的

在解决病人调度问题中不可避免。Delage和Iancu(2015)讨论了鲁棒多阶段决策产生的时间一致性问题。

**7.基于数据的效度研究**

我们进行了一个数值研究来评估混合鲁棒-随机方法的性能。使用匿名医院提供的一组患者流数据进行验证。

本院急诊科采用四级分诊制度，1级和2级急诊患者优先就诊。在这个急诊科，紧急病人被单独治疗在一个指定的区域，有专门的人员和设施。70%以上的患者属于三级。虽然他们的病情看起来稳定，但这些病人需要及时治疗以解决他们的急性症状。当急诊科人满为患时，这群病人最有可能遭受长时间等待的痛苦。1 事实上，3级患者的拥挤一直是急诊科最严重的问题，为了解决这个问题，我们将重点放在3级患者的排班上。更具体地说，我们在7.1节中给出了一些3级咨询时间数据的实证结果。将混合方法的计算性能与第7.2节中SAA方法的计算性能进行比较。我们在第7.3节中评估了动态患者调度的混合方法，其中以Huang等人(2015)提出的渐近最优调度策略为基准。

**7.1。咨询时间分类及**

**医生的异质性**

这组患者流量数据包括约12万名急诊患者的记录，其中3级患者的就诊次数超过8.5万次。每个记录包含一系列的时间戳，例如分诊、会诊和医学检查的开始和结束时间，这使我们能够重建患者在急诊科的整个路径。分诊记录和最终诊断也可以从这些记录中找到。

我们将三级患者分为两类。第一类包括最常见的急性疾病，如头痛、上呼吸道感染、急性胃炎，而第二类包括所有其他患者。在实践中，应在分诊阶段根据患者的症状和生命体征对患者进行分类。一般来说，第一类患者容易被分诊护士识别，对这些病例的诊断和治疗也相对简单。如今，越来越多的急诊科实施了一种被称为“快速通道”的项目，在这个项目中，有轻微疾病和受伤的病人由分诊护士确定，并被送到一个专门的区域进行医疗护理(Sanchez et al. 2006)。因此,病人

**何sim, Zhang:基于数据驱动的急诊科患者调度研究，2019,vol. 65, no. 4。**9, pp. 4123-4140，©2019 INFORMS

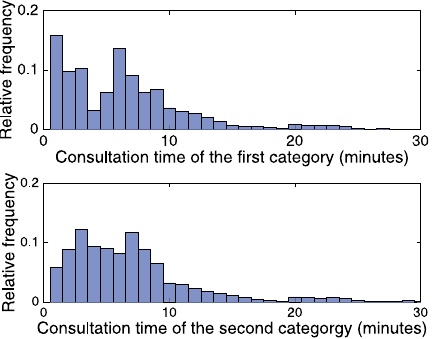
4133

分类可以很容易地通过急诊科的分诊过程来实现。我们假设当病人到达等候区时，调度系统已经知道每个病人的类别。

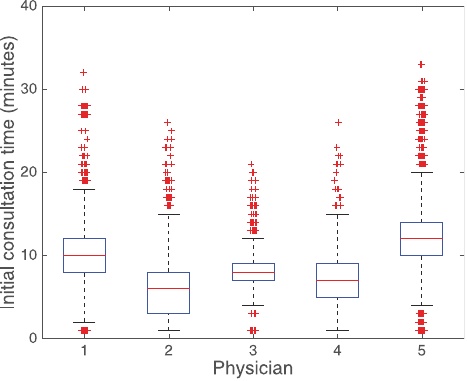
在数值研究中，我们将数据集中所有3级病例根据各自的诊断分为两组。在3级病例中，约有40%属于第一类。我们在图2中绘制了两类患者就诊时间的直方图。第一类的平均咨询时间为6.33分钟，第二类的平均咨询时间为6.94分钟。在数值实验中，假设所有医生的工作速率相同，在调度算法中使用这两个经验分布作为会诊时间的分布。由于医生可能是异质性的，每一类的经验会诊时间分布对于不同的医生可能是不同的。在这种情况下，应该使用每个医生的会诊时间记录来生成实证分布。

在这家医院的急诊科，每八小时轮班有五到六名医生为三级病人工作。尽管急诊医生被要求为各种疾病和伤害提供治疗，但他们的专业知识和工作效率各不相同。在完成了3000多个病例的医生中，我们选择了5名医生，并检查了他们所见的患者的记录。五组患者间无显著差异。我们还调查了医生在白班和夜班时的会诊时间。医师会诊时间的分布似乎并没有随时间发生很大的变化。五名医生的会诊时间箱线图如图3所示。我们可以看到这些医生的会诊时间分布有很大的不同。例如，2号医生的咨询时间中位数是5号医生的一半，而

**图2。**(彩色在线)两类咨询时间直方图



**图3。**(在线彩色)五名医生的会诊时间箱线图

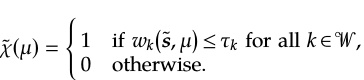


与其他医生相比，3号医生的会诊时间差异要小得多。对于一个与实践相关的病人调度方法，必须考虑到医生的异质性。

**7.2。与SAA法的比较**

容许进度得到解决(9)一般不是静态P模型的最优解问题(7)。尽管找到精确的最优解(7)是困难的,可以使用近似方法,SAA等方法,获取算法的解决方案,并减少了计算工作量。让我们比较混合方法和SAA方法的计算性能。

对于µ∈!，设χ(µ)~是由给出的指标随机变量



然后,E(˜χ(µ))?P (wk(s，µ)≤τk : k∈0)1…sN},年代n ？(snij)i∈(,j∈) n ?. .， N是咨询时间向量的一个样本，分别取自s ~和{χ1(µ)，…，χN(µ)}为χ(µ)的相应实现。根据强大数定律，在允许的时间表µ下，所有等待病人达到延迟目标的概率可以近似为 n?1Nχn(µ)/ N。利用这一事实，我们用公式表示(7)的一个近似问题



马克斯

钢桁架wk(年代n，µ)≤zn ·τk +(1−z .n)·τ¯ n，



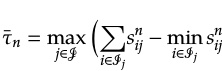
(24）

**He, Sim, and Zhang:数据驱动的急诊科患者调度**

管理科学，2019,vol. 65, no. 4。9, pp. 4123-4140，©2019 INFORMS

4134

在哪里



是病人等待时间的上限。当zn ？0，则(24)中的不等式对于所有k∈0和µ∈!都成立;当zn ？1，不等式对所有的k∈0和一个给定的µ∈!当且仅当χn(µ)?1.在给定的µ∈!下，使Nn?1zn必须等于 N n?1χn(µ)。因此，当n较大时，使(24)中的目标函数最大化的容许调度应是(7)的近似最优解。按照第5节的步骤，也可以将SAA公式转化为混合整数规划。

我们考虑一个有6名内科医生和20名急诊科患者的场景，其中8名和12名患者分别属于第一类和第二类。新增患者3例，返院患者12例，就诊患者5例。每位医生有2名返回的病人等待诊治。我们假设所有新患者都有相同的延迟目标τN 所有返回的病人都有相同的延迟目标τR．所有病人在完成当前的会诊后将离开急诊科。这六名医生的工作能力相同，因此，每次会诊时间的分配只取决于患者的类别。所有会诊时间均根据患者类别从图2的实证分布中采样。

SAA方法的计算性能主要取决于样本的大小。随着样本量的增加，我们可以得到一个更好的静态P模型问题的解决方案(7)，但代价是计算时间的延长。我们用不同的延迟目标和不同的样本量对SAA配方(24)进行了评估。对于每组参数，我们取随机样本的8种实现，然后使用每种实现求解(24)。计算时间是这一步的主要关注点。然后利用蒙特卡罗模拟，从经验分布中重新采样咨询时间，对所得到的可接受时间表进行评估。在仿真中，医生根据所获得的许可时间表对患者进行治疗。通过1000次独立的模拟运行，将所有患者达到其延迟目标的概率作为性能度量计算出来。

我们描绘SAA的表现方法在图4中,SAA (N)表示的最优解(24)基于实现的样本大小N .因为SAA方法的计算时间似乎会增加与样本容量成倍增长,我们使用图4中计算时间的对数刻度。取延迟目标对为(τN ,τR) ？(40, 50)和(35,50)分钟，我们用SAA配方进行试验

样本大小N ?20 40 100 200。如果样本量很小，解决方案在不同的实现中表现出很大的变化性，而且大多数都不是令人满意的。增加样本容量可以稳定和提高解的性能，但代价是增加计算时间。在求解SAA公式的混合整数规划时，当延迟目标较小时，计算时间有增加的趋势。当延迟目标对为(τN ,τR) ？(30,50)和(30,45)分钟，N ?200.相反，我们用N来检验这两种情况?20 40 100和150。当样本容量为N ?150，得到(τ)的最优解需要6个小时N,τR) ？30、45分钟。

我们还演示了鲁棒-随机混合方法的性能，如图4中用HRS表示。混合公式的解决方案不依赖于随机样本的具体实现，因此，在这种方法的性能中不存在可变性。虽然(9)中的目标函数与静态P模型问题(7)的目标函数不同，但混合公式的最优解优于SAA实现的大多数解。在图4中，只有当样本量较大时，SAA解决方案与HRS解决方案达成目标的概率是相当的。尽管有几种SAA实现产生比混合方法更好的解决方案，但在目标实现的概率方面，没有SAA解决方案比相应的HRS解决方案的性能高出2%以上。SAA方法需要更长的计算时间。要获得类似的性能，可能需要几个小时才能从SAA实现中获得解决方案，而获得HRS解决方案只需要几十秒。当样本量较大时，SAA方法不能用于实时患者调度。因此，我们将使用Huang et al.(2015)提出的计算效率更高的调度策略作为动态患者调度的基准策略。

**7.3。动态病人调度的实验现在让我们评估混合方法在动态病人调度中的性能。**我们关注两个绩效指标:上门服务时间超过安全限制的患者百分比，以及住院时间超过强制目标的患者百分比。由于主要的担忧是3级患者拥挤，我们将所有患者的门到提供者时间的安全限制设置为D ?30分钟，并将所有患者住院时间的强制性目标设定为K ?200分钟。在下面的数值实验中，我们根据每小时15.2个病人的泊松过程生成了5000个病人到达急诊科的流。为了便于模拟，我们

**何sim, Zhang:基于数据驱动的急诊科患者调度研究，2019,vol. 65, no. 4。**9, pp. 4123-4140，©2019 INFORMS

4135

**图4。**不同延迟目标下混合方法和SAA方法的计算性能

假设护士做的所有医疗检查和治疗都不费时间。因此，如果患者返回，他将在当前会诊结束后立即加入同一位医生的队列。急诊科共有4名医生，所有患者的会诊时间均按患者类别从图2的实证分布中抽取，其中第一类和第二类患者分别占40%和60%。

根据医院的患者流量数据，75%左右的患者在离开急诊科之前至少返回一次医生，少于4%的患者在离开急诊科之前至少返回一次医生。在数值实验中，我们假设患者最多可以返回3次，0到3次返回的概率分别为0.25、0.40、0.25和0.10。退货的数量被假定独立于病人的类别。在这些参数下，ED的负荷较大，交通强度为93.57%。

为了顺序地求解混合整数规划(23)，我们需要在触发决策迭代时指定返回患者的强制限制。因为所有的病人对上门服务的时间和住院时间有相同的要求，我们确定了三个

时间限制(T1T2T3)，在病人抵达后至他开始第二次、第三次和第四次会诊期间。这些限制的选择将影响调度算法的性能。一方面，增加这些限制将容纳更多的患者在目前的就诊延迟目标内返回，从而允许更多的新患者满足上门就诊时间的安全限制。另一方面，有了更大的强制性限制，多次咨询的患者将更有可能超过他们的住院时间目标。为了解决这两个问题，我们使用一种简单的启发式方法来确定返回患者的强制限制。由于大多数患者至少会返诊一次，因此可以通过动态调整T1，即首次返诊的强制限制，来控制新返诊与返诊之间的平衡。更具体地说，当决策迭代被触发时，我们取

| 1 |

T1吗?Tl + (tu−Tl)| 1| + |5(1)|， (25)

TL和TU是两个给定的正数，TL < TU， |1|是等待入院的新患者人数

**He, Sim, and Zhang:数据驱动的急诊科患者调度**

管理科学，2019,vol. 65, no. 4。9, pp. 4123-4140，©2019 INFORMS

4136

|是等待第二次会诊的病人人数。如果| 1 | ?5 (1) | | ?0，我们设T1 ?TL.在(25)之后，当新患者比首次就诊的患者多时，我们将提高患者第二次会诊的强制性限制，这将允许更多的新患者在上门就诊的时间限制内得到服务。当首次回国患者较多时，我们将减少T1，使更多的回国患者在其住院时间目标内得到服务。在实践中，我们认为TL和TU的时间都比上门服务的安全限制时间长几倍，这样大多数新患者就可以快速就诊。T1确定后，我们指定两个正数Δ1 和Δ2 并设置T2 ？T1 +Δ1 和T3 ？T2 +Δ2．为了便于实现，我们假设Δ1 和Δ2 在所有迭代中都是不变的。我们要求第四次会诊的强制时限T3，要远远低于住院时间限制。

在第一个数值例子中，我们假设四位医生具有相同的工作能力，因此，一个病人的会诊时间的分配不取决于具体的医生。我们使用Δ来评估这种混合方法1 ？Δ2 ？30分钟，报告不同对(TLTU)。正如大家所期望的，增加TL TU一般会减少违反“门到供应商”规定的情况，但会导致更长的滞留时间。我们比较了混合调度方法与下列调度策略的性能。

**全球先。**预约系统会在医生的咨询结束后，将预约时间最早的患者发送给该医生。由于医疗检查和治疗都假定是即时进行的，因此医生将一直为每位患者工作，直到该患者的所有咨询都完成。在这种情况下，全局FCFS策略等同于返回患者优先策略。

**新病人。**当医生完成会诊时，排班系统将有最早注册时间的新患者分配给医生;如果没有新患者，排班系统会将符合条件且挂号时间最早的归队患者分配给医生就诊。

**基准的政策。**该政策由Huang et al.(2015)提出。当医生在时间t完成会诊时，排班系统将首先检查是否有新患者的等待时间即将超过或已经超过了上门服务的时间限制(即，如果有新患者的等待时间超过或已经超过了上门服务的时间限制)。，如果i∈1，则D−(t−ti< ?，在哪里?> 0是给定的多余时间)。该排班系统将在发现新患者时优先安排，否则将优先安排返回的患者。如果要服务一个新病人，调度系统将把最早到达的那个送到可用的医生那里。[Huang et al.(2015)]对新患者采用最短期限优先政策，当所有患者都有相同的上门服务限制时，该政策被缩减为FCFS政策。[如要接待归国病人，排班系统会在预计剩余会诊时间最短的归国病人中，选择挂号时间最早的那位。换句话说，最接近完成的返回的病人将被送到可用的医生那里。[Huang et al.(2015)对返回的患者采用了广义cµ规则，以最小化累积充血成本。如果病人的住院时间被视为他的拥堵成本，他们的政策就会减少到最接近先出院的政策。在医生是同构的假设下，该调度策略被证明是渐近最优的，以最小化平均住院时间，并约束上门到提供者的时间。

在上述三种调度策略中，我们假设，当一个新病人到达等候区，发现有一个或多个医生可用时，调度

**表1。**不同调度政策下，门到供应商的时间(V)¯¯和停留时间(L)的绩效比较，每小时到达率为15.2名患者和4名同质医生

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 调度策略 | V¯ | L¯ | 3 .你的手是什么 | % l¯> 200 | %侵犯 |
| 全球先  新病人第一  基准?？1  基准?？3.  基准?？6  基准?？10  SAA (40)  小时,TL ？90年,TU ？120  小时,TL ？100年,TU ？130  小时,TL ？105年,TU ？130 | 37.64  3.87  20.59  19.63  18.39  16.39  18.66  13.31  13.59  13.40 | 52.17  74.96  59.64  59.91  61.63  62.08  61.27  63.67  63.55  64.31 | 45.88  0．10  36.52  24.34  13.66  5.80  22.12  12.80  12.10  12.54 | 0.46  8.50  4.46  4.90  6.48  6.64  4.96  2.28  2.48  2.44 | 45.88  8.50  38.28  25.76  15.92  8.35  24.08  13.15  13.09  13.26 |

**何sim, Zhang:基于数据驱动的急诊科患者调度研究，2019,vol. 65, no. 4。**9, pp. 4123-4140，©2019 INFORMS

4137

系统将随机选择一名免费的医生，患者将立即被送到该医生那里。

**南非航空公司的方法。**当一个医生完成一次会诊，或者当一个新病人到达等候区，找到至少一个免费的医生时，调度系统将解决(24)，以确定下一个病人要为可用的医生看病。

表1中我们比较几个性能措施,包括平均door-to-provider时间(¯表示为V),停留的平均时间(用L),¯door-to-provider的百分比乘以超过30分钟,持续的时间超过200分钟的比例,和侵犯患者的经验时间百分比(即。如上门服务时间超过30分钟或住院时间超过200分钟)。全球FCFS政策优先考虑归国患者，缩短了住院时间，但牺牲了就医时间。如果采用“新患者优先”政策，上门治疗的时间就会缩短，而住院的时间就会长得多。当急症室拥挤时，这两种政策都不能用于急症室以满足严格的时间限制。尽管在新的患者优先政策下，时间违规的患者比例相对较低，但此类患者的住院时间非常长，导致急诊科严重拥堵。基准政策能够在这些性能指标之间取得平衡。正如我们前面讨论过的，该策略只依赖于咨询时间的初始矩。当超出的时间被计算时，我们将报告保单的执行情况。？1 3 6或10分钟。更多的额外时间允许医生在上门治疗的时间限制内看到更多的新患者，同时带来更多的住院违规时间。随着超时时间的增加，基准策略的性能变得越来越类似于新患者优先策略。Huang等人(2015)建议?应该比门到供应商时间的安全限制小一个数量级。例如，D ?30分钟，我们可以?？3分钟是一个实用的选择。该方法的样本量为40，总体计算时间与混合方法相当。虽然SAA方法比基准策略需要更长的计算时间，但其性能仅可与?？3分钟。由于该方法没有显示出明显的优点，因此在后续的数值实验中将不采用SAA方法。

表1中的HRS表示，在所有情况下，鲁棒-随机混合方法在平均上门到供应商时间和违规停留时间百分比方面都优于基准策略。这也导致了更大比例的病人会面

门到提供商的要求，基准策略何时生效?？1、3或6分钟。请注意，当多余的时间很大时(例如，?？(6或10分钟)，在基准政策下达到住院时间目标的患者比例不令人满意。

混合方法在上门咨询时间和住院时间之间实现了更好的平衡，因为它可以评估整个咨询时间分布的影响，而不仅仅是最初时刻的影响。虽然这一优势在较高的计算成本下获得，但在实际中求解调度问题仍然是有效的。在表1中，根据我们的方法，平均逗留时间比根据基准政策略长。这是因为，在门到供应商的时间限制下，基准策略在平均停留时间方面是渐进最优的。我们的方法旨在遵守住院时间的强制性目标，出于安全考虑，达到目标的患者比例通常被视为一个更重要的性能指标。

混合方法最重要的优点是能够在异构医生在场的情况下安排病人的时间。当医生有不同的工作速率时，他们的专业知识应该被考虑到制定时间表的决定。考虑以下场景。医生是治疗第一类患者的专家，但不熟悉第二类病例。当医生有空时，在第2类中有一名新患者，其从门到提供者的时间即将超过安全限制。在那个时候，我们应该把二类病人送到这个“慢”医生那里，还是让二类病人等待，然后把一类病人送去，这样医生就可以“快”地工作?在这种情况下，必须在防止立即违反时间和减少未来拥挤之间做出权衡。需要调度策略来评估这两个操作的结果。不幸的是，基准策略不允许不同类型的医生。当医生有不同的专业知识时，减少平均住院时间不再是渐进最优的。

在第二个例子中，假设这四名医生是异质性的。其中2人擅长治疗第一类病例，但不擅长治疗第二类病例;另外两名医生对第二类医生比较有经验，但对第一类医生不熟悉。在模拟中，我们根据患者类别使用经验分布生成原始的会诊时间，而患者的实际会诊时间取决于特定的医生。如果医生是患者类别的专家，实际会诊时间为原会诊时间的80%;否则，实际咨询时间为原咨询时间的120%。所有其他模拟设置与前面的示例相同。除了

**He, Sim, and Zhang:数据驱动的急诊科患者调度**

管理科学，2019,vol. 65, no. 4。9, pp. 4123-4140，©2019 INFORMS

4138

对于前面提到的调度策略，我们还考虑以下从基准策略修改而来的调度策略。2

**修改后的基准策略。**在决定是否为新患者服务时，调度系统遵循基准策略。即当医生在时间t完成会诊时，排班系统会检查是否存在i∈1，使得D - (t - t)i< ?，在哪里?> 0是给定的多余时间。如果发现新患者，将优先考虑，否则将优先考虑归国患者。如果要为返回的患者提供服务，则调度系统遵循基准策略，选择最接近完成的患者。如果要接待一位新病人，就应该考虑到医生的专业知识。假设医生是第一类的专家。如果注册时间最早的新患者也是第一类患者，排班系统会将该患者直接发送给医生。如果该患者属于第二类，则应根据会诊时间过长造成的效率损失来评估为该患者提供服务的紧迫性。k ?1 2，让Ck 是由该医生提供的k类患者的预期总会诊时间(包括可能的未来回报)。那么，C2−C1是医生为第2类而不是第1类患者服务的额外时间。令τ(k)为第k类中最早的新患者超过门-提供商限制的时间(τ(1) ?D如果没有新的第一类患者等待就诊)。此时，τ(1)−τ(2)可用于度量第二类服务的相对紧急性，而非第一类服务的相对紧急性。由于注册时间最早的新患者属于第2类，所以我们必须让τ(1)≥τ(2)。显然，当τ(1)−τ(2)较大时，服务于类别2更为迫切。的

调度系统将通过将τ(1)−τ(2)与一个阈值进行比较来确定要服务的类别，该阈值与医生为类别2中的患者服务所需的额外时间成正比。更具体地说，当τ(1)−τ(2)≤δ(C2 −C1)对于某些δ > 0，第1类患者应尽早就诊;否则，第2类患者应最早发现。注意，我们可能有C2 −C1 ≤0，表示慢速医生也能快速为2类患者服务。在这种情况下，医生总是会为注册时间最早的新患者服务。这是合理的，因为众所周知，给予病人以短的会诊时间将减轻堵塞。如果医生是第2类的专家，调度系统可以遵循类似的程序，使用相同的系数δ确定下一个要就诊的病人。这种启发式也可以扩展到两个以上的病人类别。

在修改后的基准策略下，调度系统只有在病例相对紧急的情况下才会将患者分配给速度较慢的医生，从而更好地利用每个医生的专业知识。当医生是异质性的时候，这种简单的启发式可能会导致相当大的性能改进。在表2中，全局FCFS、新患者优先、基准政策在做出排班决策时并没有区分医生。在这些策略下，医生的平均工作效率与前面的示例相同，因此，系统的性能没有显示出很大的差异。我们评估修改后的基准政策为?？3 6和10分钟。在决定是否为新患者提供服务时遵循同样的规则，修改后的策略在上门服务时间方面并不比基准策略表现更好。相反，它的优势在于，通过让医生以更有效的方式为患者服务，缩短了住院时间。我们

**表2。**不同调度政策下，到达率为15.2名患者/小时，4名异质性医生的上门服务时间(V)¯¯和停留时间(L)的绩效比较

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 调度策略 | V¯ | L¯ | 3 .你的手是什么 | % l¯> 200 | %侵犯 |
| 全球先  新病人第一  基准?？1  基准?？3.  基准?？6  基准?？10  基准,进行修改?？3,δ?0．5  基准,进行修改?？3,δ?1  基准,进行修改?？3,δ?2  基准,进行修改?？6,δ?1  基准,进行修改?？10,δ?1  小时,TL ？90年,TU ？120  小时,TL ？100年,TU ？130  小时,TL ？105年,TU ？130 | 38.69  3.88  20.74  20.06  18.65  16.39  21.36  21.16  21.21  20.04  17.67  11.63  9.99  9.42 | 53.19  72.19  64.62  63.32  63.00  63.12  50.24  49.14  48.98  49.11  49.06  56.68  56.70  54.23 | 46.60  0．12  36.66  25.84  14.38  11.54  28.62  29.46  31.92  24.18  18.92  10.14  6.76  6.68 | 0.36  8.14  6.06  6.20  6.24  7.04  1.32  1.18  1.34  1.38  2.12  1.82  1.30  1.16 | 46.60  8.14  38.78  28.92  18.51  13.75  28.93  29.46  31.92  24.42  18.92  10.76  6.76  6.72 |

**何sim, Zhang:基于数据驱动的急诊科患者调度研究，2019,vol. 65, no. 4。**9, pp. 4123-4140，©2019 INFORMS

4139

测试δ?0.5, 1，和2 ?？三分钟，然后我们测试δ ?1对?？6分钟和10分钟。随着阈值的增大，修改后的政策将使更多的患者接受快速治疗，缩短平均住院时间。然而，随着δ的增加，修改后的政策也可能推迟更多的新患者的会诊，尽管他们的门到医生的时间即将超过安全限制。因此，增加δ既不能缩短上门服务的平均时间，也不能减少上门服务的违规行为。可以通过增加?来改善上门服务时间的性能，但这可能会降低停留时间的性能。

与前面的示例一样，混合方法可以在遵守强制性目标方面实现更平衡的性能。在门到供应商的时间方面，它优于修改的基准政策，同时保持了相当水平的停留违规时间。我们认为，这种优势还来自于评估整个分布的影响的能力。使用混合方法，调度系统可以将患者以更合适的时间快速送到医生那里，从而在所有性能指标之间达到理想的平衡。仿真结果表明，在合理的计算费用下，混合鲁棒-随机调度方法在较宽的参数范围内优于其他调度策略。

**8.结束语**

我们提出了一种数据驱动的方法来安排急诊科患者的时间，强制目标是患者上门服务的时间和住院时间。本文的主要贡献在于建立了病人调度问题的鲁棒-随机混合公式，从而以较低的计算费用获得了相应P模型问题的近似最优解。使用这种方法和实时患者流量数据，我们开发了一个动态调度算法，为每个可用的医生提供下一个病人的建议。我们的混合鲁棒-随机方法考虑了实际特性，并在数值实验中优于现有的调度策略。特别是，在异构医生在场的情况下进行调度的能力是这种方法的一个主要优势。未来，我们可能会在混合稳健-随机公式中加入额外的成本结构，以回答急诊科患者流量管理中出现的更多问题。

所提出的混合公式为求解具有延迟或吞吐量时间约束的随机网络的优化问题提供了一种计算易于处理的方法。此类问题通常出现在服务需求对时间敏感的医疗系统中，包括患者从急诊科转移到住院病房(Mandelbaum et al. 2012, Shi et al. 2016)、救护车部署(McLay和Mayorga

2013年，Maxwell等人，2014年，Chong等人，2016年)和健康检查(Baron等人，2017年)。类似的问题也可能出现在交通系统中，包括出租车调度(Seow et al. 2010)、电动汽车充电管理(Yilmaz and Krein 2013)以及随机需求和时间窗的车辆路径问题(Bertsimas and van Ryzin 1993, Fisher et al. 1997, Laporte et al. 2002, Jepsen et al. 2008)。

在最近的一篇论文中，Jaillet等人(2016)将提出的混合公式扩展到一类令人满意的问题，而这些问题通常是难以计算的。本文给出了一组与P模型等价的充分条件。遗憾的是，这些条件不适用于本文中的病人调度问题。如何量化不确定集超矩形逼近中最优性损失是我们未来研究的一个开放性问题。

分布鲁棒优化的最新进展也可能使我们能够将P模型问题转化为易于处理的形式。正如Hanasusanto et al.(2015)和Hanasusanto et al.(2017)所指出的，可以使用分布稳健的公式来减轻计算高维积分的棘手性(即:例如，当一个随机向量的分布属于某个模糊集时，可以通过求解线性或二次曲线程序得到该随机向量在给定多面体中的最坏情况概率)。通过这些技术，我们也可以将病人调度问题转化为混合整数规划。与混合方法相比，当分布信息或患者流量数据有限时，分布稳健的公式将特别有用。

**致谢**

作者要感谢副主编和审稿人的深思熟虑的意见和建设性的建议，使论文得到了显著的改进。他们特别感谢提出修改基准政策的匿名审查员。本材料中表达的任何意见、发现、结论或建议均为作者个人观点，不代表新加坡教育部或新加坡政府的观点。

**尾注**

**1**

4级是分配给非急诊病人的

很少的访问。我们不考虑4级患者，因为这一组没有延迟要求。

**2** 修改后的基准政策是由一位匿名的推荐人提出的。

**参考文献**

Armony M, israel S, Mandelbaum A, Marmor YN, Tseytlin Y, Yom- Tov GB(2015):基于数据的排队科学的视角。随机系统5(1):146 - 194。

**He, Sim, and Zhang:数据驱动的急诊科患者调度**

管理科学，2019,vol. 65, no. 4。9, pp. 4123-4140，©2019 INFORMS

4140

基于鲁棒优化的高维可控随机分析。数学。编程B 134(1): 23-70。

Bandi C, Bertsimas D, Youssef N(2015)鲁棒排队理论。③。研究》63(3):676 - 700。

王杰(2017)开放商店服务网络的战略闲置与动态调度:案例分析。③生产服务。管理19(1):52 - 71。

(2004)不确定线性规划的可调鲁棒解。数学。编程一个99(2):351 - 376。

(2004)稳健的价格。③。研究》(1):52 35-53。

基于鲁棒优化的排队网络性能分析。③。研究》59(2):455 - 466。

(2)多容量车辆在欧氏平面上的随机动态路径。③。研究》41(1):60 - 76。

Charnes A, Cooper WW(1963)在机会约束下优化和满足的确定性等价。③。研究》11(1):18-39。

张晓明(2016)应急医疗服务系统的车辆组合决策。③生产服务。管理18(3):347 - 360。

Connelly LG, Bair AE(2004)急诊科活动离散事件模拟:系统级运营研究的平台。学术急诊医学11(11):1177-1185。

de Bruin AM, Bekker R, van Zanten L, Koole GM(2010)使用Erlang损失模型对医院病房进行尺寸标注。安。③。研究》178(1):23-43。

de Ruiter FJCT, Brekelmans RCM, den Hertog D(2016)多个可调鲁棒解存在的影响。数学。编程一个160(1 - 2):531 - 545。

(2)多阶段决策的鲁棒性。Aleman DM, Thiele AC，编辑。运筹学革命，INFORMS运筹学教程(INFORMS, Catonsville, MD)， 20-46。

Doytchinov B, Lehoczky J, Shreve S(2001)基于最早截止日期优先排队原则的交通拥挤实时队列。安。达成。Probab。11(2):332 - 378。

关键词:车辆路径优化，时间窗，优化算法，车辆路径③。研究》45(3):488 - 492。

Gilboy N, Tanabe P, Travers D, Rosenan AM(2011)紧急程度指数(ESI):急诊科护理的分类工具，第4版(AHRQ出版物，Rockville, MD)。

Green LV, Soares J, Giglio JF, Green RA(2006)利用排队理论提高急诊部门人员配备的有效性。学术急诊医学13(1):61-68。

关键词:不确定性，概率约束规划，鲁棒性，分布鲁棒性数学。编程B 151(1): 35 - 62。

关键词:模糊联合机会约束，均值信息，离散信息，模糊联合机会约束③。研究》65(3):751 - 767。

Hoot NR, Aronsky D(2008)急诊拥挤的系统回顾:原因、影响和解决方案。安。急诊医学52(2):126 - 136。

Huang J, Carmeli B, Mandelbaum A(2015)急诊科患者流量的控制，或有截止日期和反馈的多等级队列。③。研究》63(4):892 - 909。

满足唤醒:缓解不确定性的模型。工作论文，新加坡国立大学，新加坡。

Jepsen M, Petersen B, sporrendonk S, Pisinger D(2008)子行不等式应用于带时间窗的车辆路径问题。③。研究》56(2):497 - 511。

Laporte G, Louveaux FV, van Hamme L(2002)具有随机需求的有能力车辆路径问题的整数L型算法。③。> 50(3): 415 - 423。

马格拉斯C, Van Mieghem JA(2005)具有前置时间约束的排队系统:一种允许和排序控制的流模型方法。欧元。j .打开。研究》167(1):179 - 207。

Mandelbaum A, Momˇcilovi´c P, Tseytlin Y(2012)关于从急诊部门到医院病房的公平路由:具有异构服务器的QED队列。管理科学。58(7):1273 - 1291。

(1)基于多目标优化的最优救护车调配策略研究。③。研究》62(5):1014 - 1027。

McHugh M(2013)急诊科拥挤和延误对病人的后果。Hall R, ed.《病人流量:减少医疗保健服务的延迟》，第二版(施普林格，纽约)，107-127。

McLay LA, Mayorga ME(2013)一个服务器到客户系统的调度模型，平衡效率和公平。③生产服务。管理15(2):205 - 220。

(2006)机会约束规划的凸逼近。王志强，王志强，王志强，等。2014 .中国海洋大学学报(自然科学版):306 - 310。

(2)一类具有吞吐量时间约束的多类队列的渐近最优动态控制。排队系统39(1):23-54。

Rabin E, Kocher K, McClelland M, Pines J, Hwang U, Rathlev N, Asplin B, Trueger NS, Weber E(2012)急诊室“登机”和拥挤的解决方案没有得到充分利用，可能需要立法。卫生事务31(8):1757 - 66。

Saghafian S, Hopp WJ, Van Oyen MP, Desmond JS, Kronick SL(2012)患者分流作为一种提高急诊科反应性的机制。③。> 60(5): 1080 - 1097。

Saghafian S, Hopp WJ, Van Oyen MP, Desmond JS, Kronick SL(2014)复杂性增强分诊:一种提高患者安全和操作效率的工具。③生产服务。管理16(3):329 - 345。

Sanchez M, Smally AJ, Grant RJ, Jacobs LM(2006)快速通道区域对急症部门绩效的影响。J.急诊医学31(1):117-120。

李德华(2010)一种多智能体协同出租车调度系统。IEEE反式。自动化科学。Engrg。7(3):607 - 616。

史鹏，周明明，戴建刚，丁东，沈俊(2016)医院住院病人住院时间的模型研究。管理科学,62(1):28。

Sinreich D, Marmor Y(2005)急诊科操作:开发模拟工具的基础。国际教育协会37(3):233 - 245。

基于多实例化的鲁棒软件分区。《信息科学与工程学报》。24(3):500-515。

Weber EJ, Mason S, Carter A, Hew RL(2011)清空耻辱走廊:英国4小时应急吞吐量目标的组织经验。安。急诊医学57 (2):79 - 88. - e1。

韦尔奇SJ，阿斯普林BR，斯通格里菲斯S，戴维森SJ，奥古斯丁J，舒尔J(2011)急诊部门运营度量、度量和定义:第二次绩效度量和标杆峰会的结果。安。急诊医学58(1):33-40。

Yilmaz M, Krein PT(2013)电池充电器拓扑结构、充电功率水平和基础设施的插电式电动汽车和混合动力汽车。IEEE反式。电力电子28(5):2151 - 2169。

Yom-Tov GB, Mandelbaum A (2014) Erlang-R:一个时变的、客户可重入的队列，用于支持医疗保健人员。③生产服务。管理16(2):283 - 299。

关键词:鲁棒最优控制，不确定性集合，不确定性集合，鲁棒最优控制汽车- matica 75:249 - 259。