

文章编号:1006-2467(2022)08-1067-11

DOI: 10.16183/j.cnki.jsjtu.2021.122

面向时变需求的多等级急诊患者入院控制

徐捷¹, 王子翔¹, 刘玉欣¹, 刘冉¹, 杨之涛²

(1. 上海交通大学机械与动力工程学院, 上海 200240;

2. 上海交通大学医学院附属瑞金医院急诊科, 上海 200025)

摘要: 建立患者准入控制的 Markov 决策过程(MDP)模型,并基于均匀化方法对该模型进行拓展,实现逐时段的实时决策过程.拓展经典 MDP 迭代求解方法,提出双向迭代算法、逐时段策略迭代算法等方法对模型求解.以上海某大型医院抢救室为例,数值实验表明:对比现有决策方法,所提出的新方法可以提高危重患者的接收率,提升急诊医疗的服务水平.

关键词: 患者入院控制;Markov 决策过程;均匀化方法;时变需求

中图分类号: TP 13; R 197.32

文献标志码: A

Admission Control of Multi-Level Emergency Patients with Time-Varying Demands

XU Jie¹, WANG Zixiang¹, LIU Yuxin¹, LIU Ran¹, YANG Zhitao²

(1. School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;

2. Department of Emergency, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China)

Abstract: The Markov decision process (MDP) model is developed for the patient admission control problem and is extended based on the uniformization method to realize a real-time period-by-period decision process. The classical MDP iterative method is extended, and the two-way iteration algorithm and the two-way threshold iteration algorithm are proposed to solve the new model. Numerical experiments are conducted based on the data from the resuscitation room of a large hospital in Shanghai. The results show that the proposed method can improve the admission rate of critical patients and enhance the medical service level of the hospital compared with the existing decision method.

Key words: patient admission control; Markov decision process (MDP); uniformization method; time-varying demands

急诊部拥挤、病床资源紧张的现象在大型三甲医院经常出现,其原因十分复杂.首先,急诊不能采用预约机制,在一天的不同时段患者到达速率高度时变且不确定.同时,急诊患者的病情繁杂且有轻重

缓急区别.进一步,每个患者占用医疗资源的时间具有不确定性,例如使用病床等医疗资源的时间不确定.这些复杂因素给急诊部服务管理和调度造成困难.

收稿日期:2021-06-17

基金项目:国家社会科学基金资助项目(19BGL245)

作者简介:徐捷(1997-),男,安徽省池州市人,硕士生,研究方向为医疗服务管理和随机系统建模优化.

通信作者:杨之涛,男,主任医师,电话(Tel.):021-64370045;E-mail:yangzhitao@hotmail.fr.

抢救室是急诊部的核心科室,而床位是抢救室最关键的资源,其使用的科学性与否直接影响对患者的救治是否及时、是否能做到对患者“应收尽收”的医疗原则。因此,急诊部需要在预诊时对患者病情进行分级,对病情严重的患者优先接收;对病情相对不紧急的患者根据资源占用情况选择性接收,即进行科学的入院控制。本文称前者为“危重患者”,后者为“非危重患者”。抢救室床位资源有限,如何通过合理的手段,对不同等级的患者加以控制,以提高患者健康回报和医院收益,是抢救室面临的重要问题。由于抢救室患者一般允许的等待时间很有限,所以当患者到达时需要实时决定是否接收。但是现实中做出科学实时的入院控制决策具有很大的难度,这是由于后续患者的到达、患者占用资源的时间都是高度不确定的,接收过多非危重患者导致床位不足,将可能影响后续危重患者的接收;接收过少则可能造成床位资源利用率不高,影响整体的医疗服务质量和收益。因此,抢救室需要科学的患者入院控制方法和策略,以提升整体运行和服务水平。

抢救室入院控制是医疗服务准入控制问题的典型场景之一。医疗服务准入控制是指针对待服务患者,对医疗资源(床位、检查设备等)进行动态分配调度,确定何种患者在什么时间可以获得医疗资源和服务,其主要研究方法包括 Markov 决策过程(MDP)^[1-4]、随机规划^[5]和近似动态规划^[6-7]等,其主要对象分为对非预约患者和预约患者的准入控制。首先,很多文献对非预约患者进行了服务准入控制研究。文献[1]针对突发事件后的入院控制问题,考虑时变的到达率和奖励函数,建立连续时间 MDP 模型,并通过状态离散化的方式进行求解。文献[3]考虑质子治疗场景下治疗组合的比例约束,通过聚合 MDP 模型,求解获得近似最优的患者入院策略。文献[5]为提高资源利用率和降低服务成本两方面的目标,提出了双目标随机优化模型,在不确定的需求和能力下取得资源利用率和服务成本之间的最优均衡。文献[4]针对不同紧急程度患者的入院控制问题,提出了一种配额策略,在获得需求信息之前决定最大日接纳量,并证明在特定条件下该策略等价于已知需求信息的最优策略。文献[8]考虑紧急患者和预约患者竞争的入院控制问题,证明了最优配额策略呈现单调性。文献[2]则研究患者到达及病情演变均不确定条件下的入院控制问题,建立资源动态分配的 MDP 模型,并通过粒子群算法进行求解。文献[9-11]进一步考虑了重症监护室场景,在拒绝新来患者和让患者提前出院之间取得平衡。

除了随机非预约到达患者,目前对预约患者的准入控制也有较广泛研究,主要包括对患者的预约调度和手术择期等研究。对于医疗检查资源的预约调度问题,文献[12]考虑爽约率和到达率不同的多类患者,利用分层算法框架求解。文献[7]针对择期手术决策问题,通过近似动态规划求解,仿真显示算法结果使得医疗系统运行效果得到改善。文献[13]针对择期患者入院控制问题,建立混合整数规划模型,求解多种资源、多个时间段和多类患者的准入控制问题。文献[14]针对多台设备多类患者场景的检查资源预约调度问题建立 MDP 模型,求解使得收益最大化。

本文主要采用 MDP 方法进行患者准入研究。虽然如上所述的 MDP 方法已经在相关问题得到了运用,但本研究的抢救室准入控制问题和以上文献有显著区别,针对抢救室这个具体场景下的准入控制问题,本文的研究对象和基本假设都有别与已有研究。

(1) 传统基于 MDP 的患者准入研究假设提前设定了决策时刻或时间槽长度,如每隔 10 min 决策一次,两个决策时刻之间(10 min 长度内)不做任何决策,以简化模型和求解。而本文突破以上设定,允许时段内发生多个事件且进行多次决策,即患者到达时立即根据系统状态做出相应的决策,更加符合抢救室的实际运行需要。

(2) 传统 MDP 患者准入控制模型设定系统只有在决策时刻才会发生状态变化。因此需要对系统的随机特征加以近似和限定,例如一般假设服务时间虽然是随机量,但是离散为时间槽的整数倍。本文的 MDP 模型中突破此约束,允许各个随机量具有连续随机性,更加符合实际的随机特点。

(3) 传统 MDP 模型由于设定了特定的决策时刻,为了获得好的效果必然要求决策时刻密集,且对每个决策时刻均求解出一个决策策略,因此最终得到很多决策策略。本文突破此模式,针对一个较长时段求解得到统一策略且不限决策时间,如此更好满足了急诊抢救室运作管理的需要,提高了研究成果实施可行性。

1 问题与模型描述

本文聚焦急诊抢救室床位资源,研究如何通过科学的手段,将有限的床位分配给不同等级的患者,即当患者到达抢救室时,按照何种策略决定是否接收该患者,以提高对患者的服务水平并提升医院的综合收益。针对急诊抢救室入院控制问题,本文根据

合作医院的调研情况和实际数据做出以下几点假设.

(1) 急诊室床位总数为 N , 每个床位可视为此服务系统中的“服务台”.

(2) 患者实际病情复杂多变, 根据我国卫生部 2011 年发布的急诊病人病情分级指导原则^[15], 急诊患者可分为 4 个等级, 其中需进入抢救室抢救的患者可分为两个严重等级, 本文称为危重患者和非危重患者.

(3) 两类患者到达速率的时变性质, 参考现有文献^[16], 将一个较长决策期等分为 T 个时段, 每个时段长度为 Δ . 例如合作医院的数据中 $T=24$, Δ 为 1 h.

(4) 根据合作医院实际数据统计拟合以及相关文献的分析结果^[17], 设定每个时段内每类患者的到达数服从泊松分布, 到达率在不同时段变化. 危重患者和非危重患者在时段 $t(t=1, 2, \dots, T)$ 内的到达速率参数分别为 λ_t^0 和 λ_t^1 .

(5) 根据合作医院提供的数据统计拟合, 设定两类患者的医疗服务时间, 即其占用床位的时间, 分别服从给定参数为 μ^0 和 μ^1 的指数分布.

(6) 采用合作医院的基本收治规则, 即当系统有床位空闲时, 若危重患者到达, 则必须本着“应收尽收”的原则加以接收, 分配床位; 若非危重患者到达, 则可以接收, 也可以拒绝接收. 当床位已满时, 就不再接收任何患者.

(7) 假设接收患者会产生确定的正收益, 拒绝患者则会产生相应的负收益. 接收一位危重患者和非危重患者的收益分别为 r^0 和 r^1 ; 拒绝一位危重患者和非危重患者的损失分别为 p^0 和 p^1 . 注意此处的收益并非仅指经济收益, 而是考虑了患者救治难度、患者转院风险、医院经济收益以及社会责任等因素的综合性指标.

本研究寻找科学的策略集合, 使得时变需求下患者的入院控制问题最优, 即求解每个时段的患者入院控制策略, 在每个时段内使用对应的最优策略, 以实现一个较长时域内(例如 24 h)总收益最大化. 在一个时段内患者随机连续到达, 患者每次到达需要实时决策, 这造成本问题中一个时段内虽然策略是确定的, 但是决策的时间点和次数不确定. 同时, 本文放弃了类似研究常用的“时间槽”概念, 设定患者的到达是随机且速率时变的, 每次到达实时决策, 这样的设定更加符合实际情况, 同时也更加具有挑战性.

以上决策问题可以通过有限期无折扣的 MDP

模型来描述. 本文建立的 MDP 模型主要包括 4 个元素, 即系统状态、决策集合、状态转移概率和收益评估.

1.1 系统状态

定义系统状态为 $s = (n_h, n_l, h)$, n_h, n_l 分别为当前系统中危重患者和非危重患者数量; h 为系统当前事件性质, 取值 0, 1, 2 分别表示“危重患者到达”“非危重患者到达”和“无患者到达”, 其中“无患者到达”包括“患者出院”和“系统自转移”(系统自转移见下文 1.3 节定义) 两类事件. 考虑到系统人数不超过 N , 因此状态总数为

$$M = 3 \sum_{i=0}^N (N-i+1) = \frac{3}{2}(N+1)(N+2) \quad (1)$$

1.2 决策集合

考虑患者到达时的决策包括“接收”和“拒绝”. 根据本文假设, 在某些状态下, 其对应决策集只有一个决策, 如床位占满时, 只能拒绝患者; 有空床且危重患者到达时, 只能接收. 需要指出, 无患者到达时, 无需进行决策, 即定义为“空决策”, 不产生收益或损失. 综上, 定义决策集为 $A = \{0, 1, 2\}$, 其中 0 表示拒绝患者, 1 表示接收患者, 2 表示空决策. 若一个时段 t 内各个状态对应的决策均确定, 则称该时段策略 π_t 确定, 任意状态 s 对应决策 $a_t(s)$ 可由该策略给出, 即对任意 s 有 $a_t(s) = \pi_t(s)$.

1.3 状态转移概率

相关 MDP 文献一般是将决策期划分为多个等长的时间槽, 假设事件发生的时间间隔是离散随机, 即为时间槽的整倍数, 从而将模型简化为决策时刻和系统状态转移均只发生在每个时间槽端点^[1, 4]. 但由于本研究中患者到达时间和服务时间均为连续随机变量且需要实时决策, 所以, 使用均匀化方法^[18]将系统事件发生时间离散化. 对于一个连续时间马尔科夫链, 令 γ 表示其最大转移速率, 则系统在时段 t 内发生事件数量 $U(t)$ 服从参数为 $\gamma\Delta$ 的泊松分布, 例如, 发生事件数量为 n 的概率为

$$P(U(t) = n) = e^{-\gamma\Delta} \frac{(\gamma\Delta)^n}{n!} \quad (2)$$

那么, 若系统当前状态为 s_i , 发生一次事件后转移到 s_j 的概率为

$$P_{ij} = \begin{cases} v_{ij}/\gamma, & s_i \neq s_j \\ 1 - \sum_{i \neq j} (v_{ij}/\gamma), & s_i = s_j \end{cases} \quad (3)$$

式中: v_{ij} 为系统从状态 s_i 到状态 s_j 的转移速率. 当 $s_i = s_j$ 时, 表示状态不发生改变, 即系统自转移.

针对本文研究的系统, 其最大转移速率 γ 可定

义为

$$\gamma = \max_{1 \leq t \leq T} (\lambda_t^0 + \lambda_t^1 + N \max\{\mu^0, \mu^1\}) \tag{4}$$

假设转移前系统状态为 $s = (n_h, n_l, h)$, 采用的决策为 $a \in A$; 决策后的瞬时状态为 s^a ; 转移后状态为 $s' = (n'_h, n'_l, h')$. 则系统在时段 t 状态转移概率 $P_t(s' | s, a)$ 可讨论如下.

场景 1 危重患者到达且接收, 即 $h=0, a=1$.
该决策的条件为 $n_h + n_l < N$. 接收患者后, $s^a = (n_h + 1, n_l, 2)$. 接收后, 可能发生的随机事件如下.

(1) 患者到达.
转移后状态为 $s' = (n_h + 1, n_l, h')$, 根据到达的患者类型 h' 取值为 0 或 1, 相应的转移概率分别为 λ_t^0/γ 和 λ_t^1/γ .

(2) 患者出院.
若危重患者出院, 转移后状态为 $s' = (n_h, n_l, 2)$, 转移概率为 $(n_h + 1)\mu^0/\gamma$; 若非危重患者出院, $s' = (n_h + 1, n_l - 1, 2)$, 转移概率为 $n_l\mu^1/\gamma$.

(3) 自转移.
状态保持不变, 即 $s' = s^a$, 转移概率为 $1 - (\lambda_t^0 + \lambda_t^1 + (n_h + 1)\mu^0 + n_l\mu^1)/\gamma$.

综上, 该场景下状态转移概率可由下式表示:
 $P_t(s' | s, a) =$

$$\begin{cases} \lambda_t^h/\gamma, & n'_h = n_h + 1, n'_l = n_l, h' \in \{0, 1\} \\ (n_h + 1)\mu^0/\gamma, & n'_h = n_h, n'_l = n_l, h' = 2 \\ n_l\mu^1/\gamma, & n'_h = n_h + 1, n'_l = n_l - 1, h' = 2 \\ 1 - [\lambda_t^0 + \lambda_t^1 + (n_h + 1)\mu^0 + n_l\mu^1]/\gamma, & n'_h = n_h + 1, n'_l = n_l, h' = 2 \end{cases} \tag{5}$$

场景 2 非危重患者到达且接收, 即 $h=1, a=1$.

该决策的条件为 $n_h + n_l < N$. 接收患者后, $s^a = (n_h, n_l + 1, 2)$. 接收后, 同上分析可得该条件下的状态转移概率为

$P_t(s' | s, a) =$

$$\begin{cases} \lambda_t^h/\gamma, & n'_h = n_h, n'_l = n_l + 1, h' \in \{0, 1\} \\ n_h\mu^0/\gamma, & n'_h = n_h - 1, n'_l = n_l + 1, h' = 2 \\ (n_l + 1)\mu^1/\gamma, & n'_h = n_h, n'_l = n_l, h' = 2 \\ 1 - [\lambda_t^0 + \lambda_t^1 + n_h\mu^0 + (n_l + 1)\mu^1]/\gamma, & n'_h = n_h, n'_l = n_l + 1, h' = 2 \end{cases} \tag{6}$$

场景 3 患者到达, 决策为拒绝或无患者到达, 即 $a=0$ 或 $h=2$.

对于危重患者到达, 该决策的条件为 $n_h + n_l = N$; 对非危重患者, 决策条件为 $n_h + n_l \leq N$; 无患者

到达时, 采取空决策. 决策后状态均为 $s^a = (n_h, n_l, 2)$. 拒绝后, 同上分析可得该条件下的状态转移概率为

$P_t(s' | s, a) =$

$$\begin{cases} \lambda_t^h/\gamma, & n'_h = n_h, n'_l = n_l, h' \in \{0, 1\} \\ n_h\mu^0/\gamma, & n'_h = n_h - 1, n'_l = n_l, h' = 2 \\ n_l\mu^1/\gamma, & n'_h = n_h, n'_l = n_l - 1, h' = 2 \\ 1 - (\lambda_t^0 + \lambda_t^1 + n_h\mu^0 + n_l\mu^1)/\gamma, & n'_h = n_h, n'_l = n_l, h' = 2 \end{cases} \tag{7}$$

场景 4 对其他未讨论情况, 有 $P_t(s' | s, a) = 0$.

以上为发生一次事件时的单步状态转移, 而当时段 t 内策略 π_t 确定, 即每个状态 s 对应的决策 a 随之确定, 场景 1 ~ 4 的单步状态转移可表示为二维状态转移矩阵 P_t ; 给定 t 时段初状态分布 (即处于各状态的概率) $\omega_t = [\omega_{t,1} \ \omega_{t,2} \ \cdots \ \omega_{t,M}]$ 后, 若该时段发生事件数为 n , 则时段末状态分布为 $\omega_t (P_t)^n$, 而根据均匀化方法, 发生事件数服从泊松分布, 由此可利用函数 $D_t(\omega_t, \pi_t)$ 计算 t 时段末的系统状态如下:

$$\omega_{t+1} = D_t(\omega_t, \pi_t) = \omega_t \sum_{n=0}^{\infty} (P_t)^n e^{-\gamma\Delta} \frac{(\gamma\Delta)^n}{n!} \tag{8}$$

在系统状态 ω_{t+1} 中, 状态 s_i 对应的概率 $\omega_{t+1,i}$ 记为 $Q_t(s_i | \omega_t, \pi_t)$, 即时段状态转移概率, 表示 t 时段初系统状态分布为 ω_t , 采用策略 π_t , $t+1$ 时段初处于状态 s_i 的概率. 由于当事件数 n 较大时, 其发生概率 $e^{-\gamma\Delta} (\gamma\Delta)^n/n!$ 接近于 0, 在数值实验中将其截断, 给定事件数上限 U_{\max} , 从而只对 $n \in [0, U_{\max}]$ 求和即可, 下文类似.

此处引入状态分布是必要的, 一方面是均匀化方法的需要, 另一方面, 本文目标为求解每个时段的患者入院控制策略, 即该时段每个状态的最优行动. 经典的 MDP 仅需要分别确定每个状态的最优行动, 不需要同时考虑其他状态. 然而本文场景下, 时段内决策次数不确定, 在时段内可能转移到其他任何状态. 如果分别从每个状态出发计算策略, 则可能在不同的状态下得到不同的策略, 这与本文要求冲突. 因此, 需要在时段开始设定状态的分布, 利用分布计算并得到时段的唯一最优策略.

1.4 收益评估

经典 MDP 模型为每个决策时刻确定最优策略, 不同决策时刻策略往往不同. 本文是为每个时段确定最优策略, 即本时段内每当患者到达抢救室, 均采用此策略决策, 最终实现决策期 (T 个时段) 内总的收益最大化. 由于每个时段内的决策时刻和决策

次数是不确定的,因此,本文在决策时所考虑的收益也区别于经典 MDP,经典 MDP 考虑一次决策后获得的“单步收益”,而本文需要考虑一个时段内“多次决策的总收益”.本文通过均匀化方法进行收益评估.

首先考虑单步收益,即在某一状态下做一次决策所能获得的收益.对于同类型患者,接收或拒绝的收益是确定而唯一的.若时段 t 状态 $s = (n_h, n_l, h)$ 采取的决策为 a ,则一次决策后的单步收益可以记为

$$c_{t,s} = \begin{cases} r^h a + p^h (1-a), & h = 0, 1 \\ 0, & h = 2 \end{cases} \quad (9)$$

文献中一般是将决策期划分为多个固定长度的时间槽,假设事件的决策时刻均在时间槽端点,进而最大化逐点的收益之和^[1].在本文研究的场景下,系统事件发生的时间间隔是连续随机的,不一定为时间槽的倍数,因此需要对收益评估做出调整.给定 t 时段初状态分布 ω_t ,当该时段策略 π_t 确定,单步状态转移矩阵 P_t 确定,由式(9)可知,每个状态在决策后能够获得的单步收益也随之确定,记为向量 $c_t = [c_{t,1} \ c_{t,2} \ \cdots \ c_{t,M}]^T$.若时段 t 内系统未发生转移,则收益为 0;否则,系统发生第 $i (i > 0)$ 次转移后的收益可表示为 $\omega_t (P_t)^{i-1} c_t$,则根据均匀化,时段 t 内的总收益 R_t 可通过函数 $R_t(\omega_t, \pi_t)$ 计算如下:

$$R_t = R_t(\omega_t, \pi_t) = \omega_t \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} (P_t)^i c_t e^{-\gamma \Delta} \frac{(\gamma \Delta)^n}{n!} \quad (10)$$

1.5 经典有限期 MDP 模型对比分析

以上构建的 MDP 模型与经典有限期 MDP 患者准入控制模型存在着显著区别.① 经典 MDP 模型存在确定的决策时刻,而本模型的决策时刻为患者随机的到达时刻,更加满足实时决策的需要;② 经典 MDP 考虑相邻决策时刻之间的单步状态转移,本模型考虑逐时段之间的状态转移,且是基于均匀化计算状态分布之间的转移;③ 经典 MDP 通过对每个决策时刻收益累加计算总收益,本模型则通过均匀化累加每个时段收益来计算总收益;④ 经典 MDP 模型通过确定每个决策时刻的策略来最优化目标,而本模型通过确定每个时段的统一策略来优化系统.

2 算法设计

首先从经典有限期 MDP 的 Bellman 最优性方程引入本文计算方法:

$$V_t(s) = \max_{a \in A} \left\{ C_t(s, a) + \sum_{i=1}^M p_t(s_i | s, a) V_{t+1}(s_i) \right\}, \quad \forall s \quad (11)$$

式中: $V_t(s)$ 为决策时刻 t 状态 s 的最优价值,即在 t 时刻从状态 s 出发,按最优策略决策,直到决策期结束时所能获得的总收益.由式(11)可知,为了最大化状态价值,需要综合考虑当前单步收益 $C_t(s, a)$ 和未来期望收益,当前状态和未来状态通过状态转移概率 $P_t(s' | s, a)$ 联系.

本文对状态最优价值做出扩展,需要注意的是,本文中 t 不表示第 t 个决策时刻,而表示第 t 个时段.令 $V_t(\omega_t)$ 表示 t 时段初状态分布 ω_t 的最优价值,即在 t 时段初从状态分布 ω_t 出发,按最优策略 π_t^* 决策,直到决策期结束时所能获得的总收益.特别地, $V_t(s)$ 表示在 t 时段初从状态 s 出发,按实际状态分布 ω_t 确定的最优策略 π_t^* 决策,直到决策期结束时所能获得的总收益. $V_t(\omega_t)$ 和 $V_t(s)$ 存在下述关系:

$$V_t(\omega_t) = \omega_t \cdot \begin{bmatrix} V_t(s_1) \\ V_t(s_2) \\ \vdots \\ V_t(s_M) \end{bmatrix} \quad (12)$$

其证明是显然的:由式(8)和式(11)可知,当最优策略 π_t^* 确定后,时段状态转移和时段收益评估都是对初始状态分布的线性变换,因此状态分布价值也就为各状态价值的线性组合,其权重恰为分布中各状态的概率.

根据状态价值定义以及式(12),本文提出 Bellman 最优性方程如下式所示:

$$V_t(\omega_t) = \max_{\pi_t} (R_t(\omega_t, \pi_t) + V_{t+1}(D_t(\omega_t, \pi_t))) = \max_{\pi_t} (R_t(\omega_t, \pi_t) + \sum_{i=1}^M Q_t(s_i | \omega_t, \pi_t) V_{t+1}(s_i)) \quad (13)$$

需要注意的是,经典 Bellman 方程式(11)中是分别对每个状态做出最优决策.但是本文 Bellman 方程式(13)考虑逐时段的递推关系,一个时段内可能会发生多次状态转移.根据式(10),要得到整个时段的总收益,须确定该时段完整的策略,因此不能分别求每个状态的最优决策,而是针对时段初状态分布,直接确定时段内最优策略.

为最大化决策期的总收益,本文基于 Bellman 最优性方程式(13)设计了双向迭代算法,确定每个时段的最优策略,即得到包含 T 个策略的策略集合.由于该算法复杂度较高,无法应对大规模问题,所以进一步提出逐时段策略迭代算法.另外,为便于实际应用,设计了双向阈值迭代算法来求解最优阈值策略.

2.1 双向迭代算法

由式(8)可知,给定决策期初始状态分布 ω_1 ,若各个时段策略确定,则之后各个时段初的状态分布 $\omega_t(t=2,3,\cdots,T)$ 均可确定;再由式(13)结合 $V_t(s)$ 定义,可从时段 T 向时段 1 方向依次计算各时段各个状态的价值 $V_t(s)$ ($t=T,T-1,\cdots,1$). 经典有限期 MDP 采用基于 Bellman 方程的逆向迭代求解,但基于本文的 Bellman 方程式(13)无法实现这样的求解过程,其原因在于未知上一时段初的状态分布 ω_t ,从而无法评估时段内的收益 R_t 和时段状态转移概率 $Q_t(s_i|\omega_t,\pi_t)$;同时,若采用正向求解,也会遇到未知下一时段初状态价值的困难.因此,本文设计双向迭代算法求解每个时段的最优策略,其中,正向寻优以时段 1 为起点,基于逆向寻优得到的各时段状态价值向后逐时段寻找最优策略,并更新各时段的状态分布,如图 1 所示;逆向寻优以时段 T 为起点,基于正向寻优得到的各时段的状态分布向前逐时段寻找最优策略,并更新各时段的状态价值,如图 2 所示;这个完整的过程称为一轮双向迭代.当相邻迭代中正向寻优所得策略不变时,算法收敛.

双向迭代算法步骤如算法 1 所示.其中: π_t^m 为 t 时段第 m 次迭代所得策略; m 为迭代迭代编号; $R_t(s,\pi_t)$ 中 s 对应 ω_t 的一个特例,表示处于状态 s 的概率为 1,处于其他状态的概率为 0;与之对应, $Q_t(s'|s,\pi_t)$ 是 $Q_t(s'|\omega_t,\pi_t)$ 的特例.

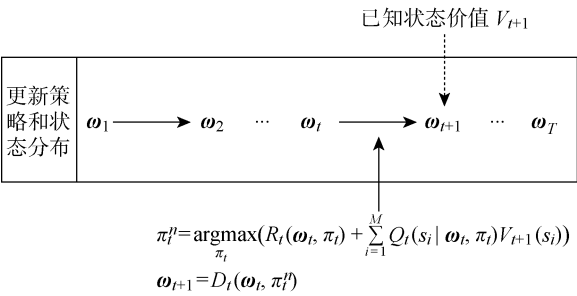


图 1 正向寻优示意图
Fig. 1 Illustration of forward search

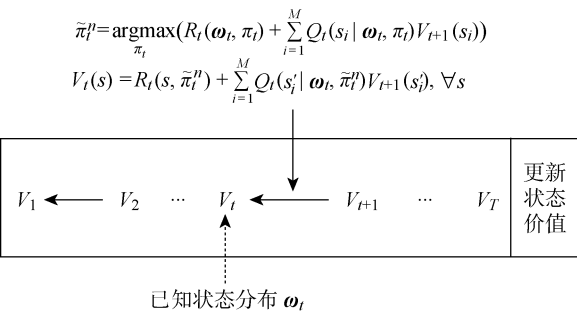


图 2 逆向寻优示意图
Fig. 2 Illustration of backward search

算法 1 双向迭代算法

```
输入  初始状态分布  $\omega_1$ 
输出  各时段策略

(1) function 双向迭代  $f_1(\omega_1)$ 
(2)    $V_t(s)=0, \forall t, s$ 
(3)   迭代编号  $m=0$ 
(4)   while  $m<2$  或  $(\pi_t^m \neq \pi_t^{m-1}, \forall t)$  do
(5)      $m=m+1$ 
(6)     for  $t=1$  to  $T$  do ▷正向寻优
(7)        $\pi_t^m = \arg \max_{\pi_t} (R_t(\omega_t, \pi_t) +$ 
 $\sum_{i=1}^M Q_t(s_i | \omega_t, \pi_t) V_{t+1}(s_i)$ )
(8)        $\omega_{t+1} = D_t(\omega_t, \pi_t^m)$ 
(9)     end for
(10)    for  $t=T$  to 1 do ▷逆向寻优
(11)       $\tilde{\pi}_t^m = \arg \max_{\pi_t} (R_t(\omega_t, \pi_t) +$ 
 $\sum_{i=1}^M Q_t(s_i | \omega_t, \pi_t) V_{t+1}(s_i)$ )
(12)       $V_t(s) = R_t(s, \tilde{\pi}_t^m) +$ 
 $\sum_{i=1}^M Q_t(s'_i | s, \tilde{\pi}_t^m) V_{t+1}(s'_i), \forall s$ 
(13)    end for
(14)  end while
(15)  return  $\pi^m = \{\pi_1^m, \pi_2^m, \cdots, \pi_T^m\}$ 
(16) end function
```

数值实验验证了该算法在小规模数据上的最优性,需要注意的是,在所有状态中,危重患者到达或无患者到达时,决策是确定的.只有当非危重患者到达时,可能的决策有两种,可得到每个时段内不同的策略共有 $2^{M/3}$ 种,与状态总数 M 呈指数关系.由于在算法 1 步骤 7 和 11 中直接遍历所有策略,显然该算法难以应用到大规模抢救室入院控制问题.

2.2 逐时段策略迭代算法

为求解大规模问题进一步提出“逐时段策略迭代算法”,求解近似最优策略.该算法从时段 T 向时段 1 依次寻优,对每个时段,采取策略迭代算法,先随机选取一个策略,如先到先服务(FCFS)策略,再逐状态改进当前策略,直到相邻迭代所得策略不变,则该时段迭代过程结束,然后继续对前一个时段进行策略迭代,直到所有时段策略确定.数值实验显示,每个时段一般不超过 4 轮迭代策略即确定,而每轮迭代需评估的策略数仅为 $2M/3$,求解效率大幅

提升.

逐时段策略迭代算法具体步骤如算法 2 所示. 其中: $Q_t(s' | \pi_t^m, s, a)$ 为设定策略 π_t^m 中状态 s 对应决策为 a , t 时段从状态 s 到状态 s' 的转移概率; $R_t(\pi_t^m, s, a)$ 为设定策略 π_t^m 中状态 s 对应决策为 a , t 时段初从状态 s 出发, 得到的总收益.

算法 2 逐时段策略迭代算法

输入 初始策略 π^{st} (如 FCFS)

输出 各时段策略

```

(1) function 逐时段策略迭代  $f_2(\pi^{\text{st}})$ 
(2)  $V_{T+1}(s) = 0, \quad \forall s$ 
(3) for  $t = T$  to 1 do
(4) 迭代编号  $m = 0$ 
(5)  $m_t^m = \pi^{\text{st}}$ 
(6) while  $m = 0$  或  $\pi_t^m \neq \pi_t^{m-1}$  do
(7)  $\pi_t^{m+1}(s) = \arg \max_{a \in A} (R_t(\pi_t^m, s, a) +$ 

$$\sum_{i=1}^M Q_t(s'_i | \pi_t^m, s, a) V_{t+1}(s'_i)), \quad \forall s$$

(8)  $m = m + 1$ 
(9) end while
(10)  $\pi_t^* = \pi_t^m$ 
(11)  $V_t(s) = R_t(s, \pi_t^m) +$ 

$$\sum_{i=1}^M Q_t(s'_i | s, \pi_t^m) V_{t+1}(s'_i), \quad \forall s$$

(12) end for
(13) return  $\pi^* = \{\pi_1^*, \pi_2^*, \dots, \pi_T^*\}$ 
(14) end function

```

2.3 双向阈值迭代算法

考虑到抢救室入院控制实际应用时的便利性, 本文设计阈值策略. 阈值策略即为每个时段提供一个阈值, 基于该阈值可确定该时段内的唯一策略. 考虑两种阈值策略, 空闲床位阈值策略和非危重患者阈值策略.

(1) 空闲床位阈值策略.

该策略基于系统中空闲床位的数量来决定是否接收非危重患者: 当空闲床位数量大于某一阈值 τ 时, 接收非危重患者; 否则不接收非危重患者.

(2) 非危重患者阈值策略.

该策略基于系统中已有非危重患者的数量来决定是否接收非危重患者: 当已有非危重患者的数量小于某一阈值 τ 时, 接收非危重患者; 否则不接收非

危重患者.

阈值策略可采用上文双向迭代框架求解, 本文称为双向阈值迭代算法. 相比于双向迭代算法, 只需将对策略的搜索调整为对阈值的遍历. 双向迭代算法每个时段要遍历 $2^{M/3}$ 个策略, 而由于阈值范围有限 ($[0, N]$), 每个时段只需要遍历 $N+1$ 个策略, 因此决策空间大大缩小, 可应用于较大规模场景.

双向阈值迭代算法具体步骤如算法 3 所示. 其中: τ_t 为 t 时段的阈值; $Q_t(s' | \omega_t, \tau_t)$ 为给定阈值策略 τ_t 后, 由状态分布 ω_t 转移到状态 s' 的概率; $R_t(\omega_t, \tau_t)$ 为从状态分布 ω_t 出发, 根据阈值策略 τ_t 得到 t 时段内的总期望收益; $D_t(\omega_t, \tau_t)$ 为从状态分布 ω_t 出发, 根据阈值策略 τ_t 得到 $t+1$ 时段初的状态分布 ω_{t+1} .

算法 3 双向阈值迭代算法

输入 初始状态分布 ω_1

输出 各时段阈值

```

(1) function 双向阈值迭代  $f_3(\omega_1)$ 
(2)  $V_t(s) = 0, \quad \forall t, s$ 
(3) 迭代编号  $m = 0$ 
(4) while  $m < 2$  或  $(\tau_t^m \neq \tau_t^{m-1}, \quad \forall t)$  do
(5)  $m = m + 1$ 
(6) for  $t = 1$  to  $T$  do ▷ 正向寻优
(7)  $\tau_t^m = \arg \max_{\tau \in [0, N]} (R_t(\omega_t, \tau) +$ 

$$\sum_{i=1}^M Q_t(s_i | \omega_t, \tau) V_{t+1}(s_i))$$

(8)  $\omega_{t+1} = D_t(\omega_t, \tau_t^m)$ 
(9) end for
(10) for  $t = T$  to 1 do ▷ 逆向寻优
(11)  $\tilde{\tau}_t^m = \arg \max_{\tau \in [0, N]} (R_t(\omega_t, \tau) +$ 

$$\sum_{i=1}^M Q_t(s_i | \omega_t, \tau) V_{t+1}(s_i))$$

(12)  $V_t(s) = R_t(s, \tilde{\tau}_t^m) +$ 

$$\sum_{i=1}^M Q_t(s'_i | s, \tilde{\tau}_t^m) V_{t+1}(s'_i), \quad \forall s$$

(13) end for
(14) end while
(15) return  $\tau^m = \{\tau_1^m, \tau_2^m, \dots, \tau_T^m\}$ 
(16) end function

```

3 数值实验

使用上海某大型三甲医院急诊部的实际运行数

据,首先利用处理后的小规模数据对双向迭代算法的最优性加以验证,再基于医院真实数据对比分析各个算法的实际性能,最后对床位数量进行灵敏度分析,为抢救室入院提供易于执行的控制策略和床位数量安排指导意见. 为使用均匀化方法,截断了式(8)和(10)中无限事件数,设每个时段内最多发生事件数为 $U_{\max}=50$,且通过实验验证了此设定可保证均匀化精度. 数值实验采用的数据见网络材料 [https://pan. baidu. com/s/1UaRkX-iXta2o4NBCgwthRw](https://pan.baidu.com/s/1UaRkX-iXta2o4NBCgwthRw) (提取码:jl48).

3.1 双向迭代算法最优性验证

理论证明双向迭代算法最优性非常困难. 但针对小规模问题可通过枚举法枚举出所有时段策略的组合,确定最优策略以及最大收益. 因此,本文将双向迭代算法与枚举法在多组实验参数下的收益结果加以对比,进行最优性验证. 由于医院原始数据规模较大,考虑缩短决策期和缩小状态空间来降低求解时间,使用医院采集数据中连续 6 h 且设置 3 个可用床位,在此基础上设置不同参数共计得到 8 个算例. 算例中统一的参数设定如表 1 所示,算例间参数区别包括各时段到达率和单步收益(或损失),具体参数数值见网络材料 SM-1 节. 求解结果如表 2 所示,由表 2 可见,双向迭代算法和枚举法的求解结果完全一致,数值结果支持双向迭代算法的最优性假设.

表 1 最优性验证确定参数

Tab. 1 Fixed parameters used in optimality validation

参数	取值
N	3
T	6
Δ	1
μ^0, μ^1	0. 4, 0. 63

表 2 最优性验证结果

Tab. 2 Results of optimality validation

算例	双向迭代算法总收益	枚举总收益
1	58. 565	58. 565
2	30. 453	30. 453
3	134. 397	134. 397
4	106. 729	106. 729
5	64. 915	64. 915
6	35. 923	35. 923
7	143. 311	143. 311
8	114. 536	114. 536

3.2 逐时段策略迭代和双向阈值迭代算法对比实验

由于双向迭代算法复杂度很高,难以应对实际场景带来的大规模准入控制问题,利用逐时段策略迭代算法求解近似最优策略,并从易于实施的角度,采用双向阈值迭代算法求解两种阈值策略. 本节以先到先服务策略为基准策略,记为 K0,分别与近似最优策略(记为 K1)以及两种阈值策略(记空闲床位阈值策略为 K2,非危重患者阈值策略为 K3)进行对比,每种策略均由仿真进行系统的性能评估,得到总收益和患者接收率指标. 采用急诊部提供的实际运行数据,考虑长度为一天 24 h 的决策期,床位数目、服务速率等参数如表 3 所示(完整参数见网络材料 SM-2 节).

表 3 对比实验参数

Tab. 3 Parameters used in comparison experiment

参数	取值
N	30
T	24
Δ	1
μ^0, μ^1	0. 1, 0. 333
r^0, r^1	20, 3
p^0, p^1	-30, -1

4 种策略收益及效率的对比结果如表 4 所示,表中显示均匀化评估所得收益、仿真评估所得收益(仿真 10^5 d)、算法求得的策略相比 K0 的收益提升(“收益提升”列)以及算法运行时间(取算法运行 5 次的平均时间). 由表 4 可知,在各个策略下,均匀化评估结果与仿真结果都十分接近,误差不超过 0. 03%,验证了均匀化方法的评估精度. 不同的策略下收益表现有显著差异,K1 取得了最高收益,相比 K0 提升 6. 96%;K2 与 K1 表现非常接近,差距不足 0. 1%;K3 相比 K0 提升了 3. 3%,表现不如 K2 策略. 但从效率上看,K2 和 K3 策略由于搜索空间较小,其求解效率远优于 K1 策略. 综合来看,K2 的求解结果和效率更具优越性.

表 4 各策略收益及效率对比

Tab. 4 Comparison of returns and efficiencies by policy

策略	均匀化评估 总收益	仿真评估 总收益	收益 提升/%	算法运行 时间/s
K0	1 023. 18	1 023. 38	—	—
K1	1 094. 26	1 094. 58	6. 96	15 481
K2	1 093. 97	1 094. 14	6. 91	706
K3	1 057. 48	1 057. 16	3. 30	691

除了总收益外,患者接收率也是抢救室关注的重点指标,尤其是危重患者的接收率. 本文通过仿真统计 3 项患者接收率,分别为总接收率(即不区分患者类型的接收率,记为 Y)、危重患者接收率(记为 Y_0)和非危重患者接收率(记为 Y_1). 各时段平均接收率结果如表 5 所示,分时段接收率见网络材料 SM-3 节. 相比于基准策略,本文优化后的 3 种策略表现有所差异,虽然均提高了平均危重患者接收率,但导致平均非危重患者接收率有不同程度的降低. K1 和 K2 的平均危重患者接收率由 95.6% 提升到 99.1%,提高了对危重患者的服务水平,且平均非危重患者接收率保持在 84% 以上. K3 的平均危重患者接收率尽管也提高到 98.1%,但平均非危重患者接收率降低较多不足 80%. 由此可见,即使在相同的参数下,采取不同的策略,对患者接收率仍有较大

影响,本文提出的 K1、K2 策略在保证总体接收率合理的情况下,更大程度上提高了危重患者的接收率而具有优势.

由于抢救室重点关注危重患者,重点针对每个时段的危重接收率进行分析,如图 3 所示. 基准策略 K0 在不同时段波动很大,整体接收率低,难以实现应收尽收原则. K1 有 19 个时段的接收率在 98% 以上, K2 也有 17 个时段的接收率在 98% 以上,验证了 K2 阈值策略的性能优势. 且注意到 K1 和 K2 策略在 24 个时段中接收率波动较小,服务水平稳定. K3 策略相比 K0 有所提升,但有 13 个时段的接收率在 98% 以下,难以达到医院要求. 整体来看,本文求解所得 3 种策略相比基准策略都有较大提升,其中 K3 提升较少,而 K1 和 K2 提升显著,尤其是 K2 阈值策略,既有性能优势又易于实施,优势明显.

表 5 各策略平均患者接收率对比
Tab. 5 Comparison of average admission rates by policy

K0			K1			K2			K3		
Y	Y_0	Y_1	Y	Y_0	Y_1	Y	Y_0	Y_1	Y	Y_0	Y_1
0.956	0.956	0.956	0.912	0.991	0.842	0.913	0.991	0.845	0.886	0.981	0.798

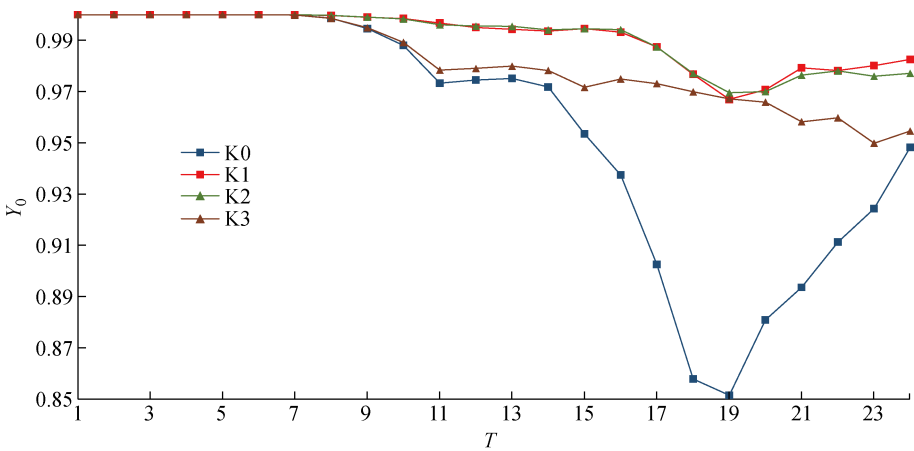


图 3 各策略逐时段危重患者接收率
Fig. 3 Admission rates of critical patients of each policy by time of day

3.3 床位数量灵敏度分析

显然,床位数量越多,医院就可以接收更多的患者,达到更高的接收率. 但是抢救室床位资源成本高昂,医护资源也有限,并不能无限扩增床位. 因此本文对床位数量进行敏感度分析,讨论不同数量的床位对危重患者接收率及总收益的影响.

除床位数量外,本节采用参数均与 3.2 节相同. 因 K2 策略结果与 K1 策略接近,且更具实际应用意义,本节采用 K2 策略进行分析,讨论在该策略下床位数量的影响. 考虑 $N \in [25, 35]$ 的变化区间,总收

益变化如表 6 所示. 由表 6 可以看出,床位增加带来收益增加,但增长速度越来越慢,即增加床位的边际收益越来越少.

除了总收益外,抢救室还关注一天内危重患者的平均接收率 R_0 . 随床位数量的变化. 不同床位数量下接收率变化如图 4 所示. 由图 4 可知,随着床位数量增长,平均危重接收率持续增长,但增长率逐渐放缓,直到增加到 30 张床位时,平均危重接收率达到抢救室目标值 $Y^* = 0.99$. 基于在合作医院调研得到的床位成本,当床位数超过 30 时,增加的总收

表 6 不同床位数量下收益变化

Tab. 6 Change in returns with different number of beds

N	均匀化评估总收益	相比 N-1 的增长率/%
25	980.51	—
26	1 010.43	3.05
27	1 036.23	2.55
28	1 058.45	2.14
29	1 077.68	1.82
30	1 093.97	1.51
31	1 107.60	1.25
32	1 118.95	1.02
33	1 128.31	0.84
34	1 135.90	0.67
35	1 142.01	0.54

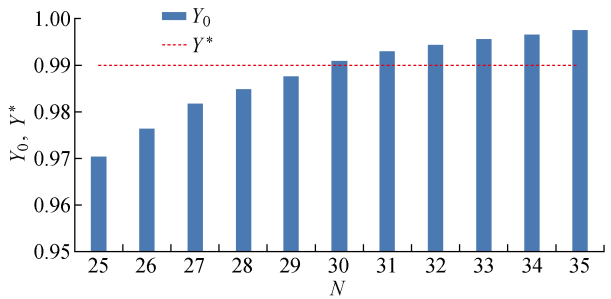


图 4 不同床位数量下接收率变化

Fig. 4 Changes in admission rates with different number of beds

益低于床位增加成本。因此,在保证平均危重接收率达到目标危重接收率的条件下,较为合理的床位数量为 30,此时既能满足危重患者服务水平的要求,又控制了总投入成本。

4 结语

针对急诊抢救室床位资源紧张的问题,提出根据患者病情严重及紧急程度选择性收治患者。建立了 MDP 模型,考虑到到达率的高度时变特性,使用均匀化方法逐时段进行离散化并求解每个时段内的最优策略。提出了求解最优策略的双向迭代算法和求解近似最优策略的逐时段策略迭代算法,实现了在较大规模数据和较长决策期场景下的应用。为了易于在实际场景中实施,进一步设计了双向阈值迭代算法,高效地为大规模实际场景求解得到简单且有效的阈值策略。数值实验验证了双向迭代算法在小规模数据上的最优性,验证了近似最优策略以及两种阈值策略的效果,所提出的阈值策略性能与近似最优策略接近且易于实施,可以为抢救室床位管

理提供有效指导。本研究方法虽可以对时变且随机患者需求等复杂条件的准入问题进行决策,但也存在一些局限。首先受限于迭代算法复杂度较高,难以应用于大规模问题,拟进一步采用深度强化学习等方法来提高求解效率。另一方面可拓展考虑对允许加床等更复杂的场景进行准入决策研究。

参考文献:

[1] LEE H R, LEE T. Markov decision process model for patient admission decision at an emergency department under a surge demand[J]. **Flexible Services and Manufacturing Journal**, 2018, 30(1/2): 98-122.

[2] 万志远, 刘勤明, 叶春明, 等. 突发事件下的医院应急资源动态分配模型研究[J]. **计算机应用研究**, 2020, 37(2): 456-459.

WAN Zhiyuan, LIU Qinming, YE Chunming, *et al.* Research on hospital emergency resource dynamic allocation model under emergencies[J]. **Application Research of Computers**, 2020, 37(2): 456-459.

[3] GEDIK R, ZHANG S F, RAINWATER C. Strategic level proton therapy patient admission planning: A Markov decision process modeling approach[J]. **Health Care Management Science**, 2017, 20(2): 286-302.

[4] DAI J J, GENG N, XIE X L. Dynamic admission quota control with controllable and uncontrollable demands and random service time[J]. **IEEE Transactions on Automatic Control**, 2021, 66(6): 2925-2932.

[5] BATISTA A, VERA J, POZO D. Multi-objective admission planning problem: A two-stage stochastic approach[J]. **Health Care Management Science**, 2020, 23(1): 51-65.

[6] HULSHOF P J H, MES M R K, BOUCHERIE R J, *et al.* Patient admission planning using approximate dynamic programming[J]. **Flexible Services and Manufacturing Journal**, 2016, 28(1/2): 30-61.

[7] DIAMANT A, MILNER J, QUERESHY F. Dynamic patient scheduling for multi-appointment health care programs[J]. **Production and Operations Management**, 2018, 27(1): 58-79.

[8] AYVAZ N, HUH W T. Allocation of hospital capacity to multiple types of patients[J]. **Journal of Revenue and Pricing Management**, 2010, 9(5): 386-398.

[9] LI X J, LIU D C, GENG N, *et al.* Optimal ICU admission control with premature discharge[J]. **IEEE Transactions on Automation Science and Engineering**, 2019, 16(1): 148-164.

[10] KIM S H, CHAN C W, OLIVARES M, *et al.* ICU

admission control: An empirical study of capacity allocation and its implication for patient outcomes[J]. **Management Science**, 2015, 61(1): 19-38.

[11] CHAN C W, FARIAS V F, BAMBOS N, *et al*. Optimizing intensive care unit discharge decisions with patient readmissions[J]. **Operations Research**, 2012, 60(6): 1323-1341.

[12] AKHAVIZADEGAN F, ANSARIFAR J, JOLAI F. A novel approach to determine a tactical and operational decision for dynamic appointment scheduling at nuclear medical center[J]. **Computers & Operations Research**, 2017, 78: 267-277.

[13] HULSHOF P J H, BOUCHERIE R J, HANS E W, *et al*. Tactical resource allocation and elective patient admission planning in care processes[J]. **Health Care Management Science**, 2013, 16(2): 152-166.

[14] 梁峰, 徐苹. 基于 MDP 和动态规划的医疗检查预约调度优化方法研究[J]. **运筹与管理**, 2020, 29(5): 17-25.

LIANG Feng, XU Ping. Appointment scheduling of medical examination based on MDP and dynamic programming[J]. **Operations Research and Management Science**, 2020, 29(5): 17-25.

[15] 中华人民共和国卫生部. 急诊病人病情分级指导原则(征求意见稿)[J]. **中华危重症医学杂志(电子版)**, 2011, 4(4): 241-243.

Ministry of Health of the People’s Republic of China. Guidelines for grading the condition of emergency patients (draft for comments) [J]. **Chinese Journal of Critical Care Medicine (Electronic Edition)**, 2011, 4(4): 241-243.

[16] LIU R, XIE X L. Physician staffing for emergency departments with time-varying demand [J]. **INFORMS Journal on Computing**, 2018, 30(3): 588-607.

[17] 文静, 耿娜, Xiaolan Xie, 等. 基于仿真的急诊室动态调度研究[J]. **工业工程与管理**, 2021, 26(3): 160-167.

WEN Jing, GENG Na, XIAOLAN Xie, *et al*. Simulation study of dynamic scheduling for emergency department [J]. **Industrial Engineering and Management**, 2021, 26(3): 160-167.

[18] PLATZMAN L K. Introduction to probability models [J]. **Journal of Quality Technology**, 1982, 14(4): 228-229.

(本文编辑:石易文)