

## 基于拉格朗日松弛的预约调度模型与算法

陶继平<sup>1</sup>, 梅枝煌<sup>1</sup>, 黄荣欢<sup>1</sup>, 林子雨<sup>2</sup>

(1. 厦门大学 自动化系, 厦门 361005; 2. 厦门大学 计算机科学系, 厦门 361005)

**摘要** 针对带有爽约的预约调度问题, 在假定未爽约病人都在相应预约段的起始点准时到达的情况下, 构建了一个以预约人数为优化变量的整数规划模型. 目标函数包括服务病人收益、病人等待费用及系统超时费用. 通过松弛各时间段剩余人数概率的关联约束, 提出了基于拉格朗日松弛的求解算法, 其松弛问题通过动态规划求解, 对偶问题通过经典的次梯度法求解. 数值实验表明, 针对小规模预约段数时, 该算法都能找到最优解; 当预约段数较大时, 算法找到的最好解整体上优于文献中已有的算法, 从而验证了算法的有效性.

**关键词** 预约调度; 过度预约; 爽约; 拉格朗日松弛算法; 动态规划

## Lagrangian relaxation based appointment scheduling model and algorithm

TAO Ji-ping<sup>1</sup>, HUANG Rong-huan<sup>1</sup>, MEI Zhi-huang<sup>1</sup>, LIN Zi-yu<sup>2</sup> 中英文作者不一致

(1. Department of Automation, Xiamen University, Xiamen 361005, China; 2. Department of Computer Science, Xiamen University, Xiamen 361005, China)

**Abstract** With the assumption that each appointment patient up in clinic arrives on time, this paper investigates the appointment scheduling problem with no-show. An integer programming model is established with the numbers of appointment patient and probability distribution of remaining patients at each slot as optimization variables. The objective function includes the benefits of serving patients, the patient waiting time costs, and system overtime expenses. By relaxing the coupled constraints about probability distribution of remaining patients at each slot, this paper proposes a Lagrangian relaxation method to solve the model with a dynamic programming algorithm solving the relaxation problem and a sub-gradient method solving the dual problem. Numerical experiments show that the algorithm can find the optimal solution for small scale problems, and that the best solution from the proposed algorithm is better than the one in the related literature for large scale problems.

**Keywords** appointment scheduling; overbooking; no-show; Lagrangian relaxation; dynamic programming

### 1 引言

近年来, 随着医疗信息化的发展以及人们医疗需求的增加, 国家对卫生事业投入的比重逐步加大, 医疗运作管理研究得到快速发展. 预约调度问题作为医疗运作管理中一个重要环节, 吸引了众多学者的研究<sup>[1-5]</sup>. 从 1952 年 Bailey<sup>[6]</sup> 的第一篇关于预约调度规则的文献起, 预约调度经过 60 多年的发展, 涌现了大量的文献. Cayirli<sup>[7]</sup> 对门诊预约调度的预约规则、设计方法、性能评估、模型描述、分析方法作了全面的综述. 由于病人的到达时间以及服务时间的多变性、病人的就诊偏好等多方面的因素, Gupta<sup>[8]</sup> 对基础医疗和手术预约系统设计的复杂性进行了归纳总结, 并指出预约调度的模型研究充满机遇. 在建模方面, 针对医疗系统中的选址与布局、能力规划、人员以及资源调度等各个环节, Bhattacharjee<sup>[9]</sup> 从优化的角度对医疗系统的文献做了详细的综述. 罗利<sup>[10]</sup> 指出, 目前国内医疗服务资源调度优化的研究成果偏少, 但对医院运作管理实践却

收稿日期: 2015-01-08

资助项目: 国家自然科学基金青年项目 (11201391, 61303004)

作者简介: 陶继平 (1980-), 男, 汉, 安徽怀宁人, 助理教授, 博士, 研究方向: 算法设计与竞争比分析, 调度理论与算法, E-mail: taojiping@xmu.edu.cn; 通信作者: 林子雨 (1978-), 男, 汉, 吉林柳河人, 助理教授, 博士, 研究方向: 数据库、数据仓库、大数据, E-mail: ziyulin@xmu.edu.cn.

有现实意义.

患者爽约 (no-show) 是影响预约系统效率的重要因素, Rust 等<sup>[21]</sup> 指出门诊的爽约率高达 30%~80%, 这使得预约系统的设计变得复杂化. 目前大部分文献考虑的带爽约的预约调度模型集中在两类: 一类在给定总预约人数下, 求取每段最优的预约人数和间隔, 以最小化病人的等待时间、医生的空闲时间或者两者的加权和等相关费用<sup>[2,4,12]</sup>. Ho<sup>[2]</sup> 考虑的是在给定的预约规则和预约人数下, 通过设定的爽约率和服务时间分布最大化病人和医生的空闲时间加权和. Tang<sup>[4]</sup> 将病人划分为一般病人和急诊病人, 考虑的是在存在爽约的情况下, 最小化病人等待时间、医生空闲时间以及系统系统加班时间的加权和, 提出了一个基于多模的局部搜索算法. Robinson<sup>[12]</sup> 做了进一步拓展, 分析了传统预约调度和开放预约两种预约模式下病人的等待时间、医生的空闲时间以及系统加班时间的加权和, 结果显示开发式预约的结果优于传统预约. 另一类预约调度模型是在给定预约段数下, 求取最优预约人数以最大化系统期望总收益<sup>[13-15]</sup>. 曹萍萍<sup>[16]</sup> 等指出第一类问题的解决是第二类问题的前提和基础. 具体到研究问题方面, Denton<sup>[17]</sup> 在 2003 年一篇具有里程碑意义的文章中, 指出最优预约人数具有圆筒状结构, 这一理论影响深远<sup>[15,18]</sup>. 针对上述第二类模型, 求解方法基本都是基于隐式效用函数的启发式或者仿真方法<sup>[15,19]</sup>, 求得的解都是近似解, 而且近似程度基本都没有理论或者仿真估计, 这也是目前求解预约调度问题的难点. LaGanga<sup>[13-14]</sup> 第一次运用分析了的方法, 显式地构造了系统总效用函数模型, 建立了一种基于过度预约的优化模型, 该模型适用于等待及超时效用函数均为凸的情形.

本文针对带有爽约的预约调度问题, 延伸了 LaGanga 等<sup>[13-14]</sup> 的过度预约建模思想, 构建了一个以预约人数为优化变量的数学规划模型, 提出了一种基于拉格朗日松弛的求解方法, 并进行了详细的仿真实验. 全文章节安排如下, 第 2 节建立了预约调度数学模型, 第 3 节提出了基于拉格朗日松弛的算法, 第 4 节进行了仿真实验, 最后一节对全文进行了总结与展望.

## 2 门诊预约调度数学模型

### 2.1 问题描述与假设

在一个预约系统中, 假定总服务时间 (session) 为  $D$ , 每个病人的服务时间确定且相同, 记为  $d$ . 将总服务时间划分为  $N$  个等间隔的预约段 (slot), 每个预约段正好服务一个病人, 即  $D = N \times d$ . 为了便于分析, 可以假设  $d = 1$ . 假设病人的爽约率 (no-show) 都相同, 记为  $\sigma$ , 且未爽约病人都在相应预约段的起始点准时到达, 这样每个预约时段到达的总病人人数服从二项分布. 相关参数定义如表 1.

### 2.2 数学模型

给定一个预约调度  $S$ , 系统总效用为:

$$f(s) = B(s) - W(s) - O(s) \quad (1)$$

#### 2.2.1 服务病人期望收益 $B(s)$

给定第  $j$  个预约段的预约人数  $a_j$ , 该段服务病人的期望收益可表示为  $\pi(1 - \sigma)a_j$ , 总的服务病人期望收益为:

$$B(s) = \sum_{j=1}^N \pi(1 - \sigma)a_j \quad (2)$$

#### 2.2.2 等待时间产生的期望费用 $W(s)$

根据表 1 的定义, 第  $j$  个预约段结束时剩余人数为  $x$  的概率为  $\gamma_j(x)$ , 第  $j$  个预约段实际到达人数为  $m$  的概率为  $\partial_j(m)$ , 记第  $j$  个预约段结束时剩余  $x$  且第  $j + 1$  个预约段实际到达人数为  $m$  的联合概率为  $\theta_{j+1}(x, m)$ , 根据两者的独立性有  $\theta_{j+1}(x, m) = \partial_{j+1}(m)\gamma_j(x)$ , 再根据实际到达人数服从二项分布, 有

$$\partial_{j+1}(m) = \binom{a_{j+1}}{m} (1 - \sigma)^m \sigma^{a_{j+1} - m}.$$

表 1 预约调度的参数定义

参数	定义
$N$	预约段数
$\sigma$	每个预约病人的爽约率
$S$	一组可行的预约调度, $S = [a_1, a_2, \dots, a_N]'$
$a_j$	每个预约段的预约人数
$\varphi(\cdot)$	系统关于超时人数的超时费用函数
$\beta(\cdot)$	每个病人关于等待时间的等待费用函数
$\pi$	服务一个病人的收益系数
$\partial_j(m)$	第 $j$ 个预约段实际到达人数为 $m$ 的概率
$\gamma_j(x)$	第 $j$ 个预约段结束时剩余人数为 $x$ 的概率
$UB_j^s$	第 $j$ 个预约段预约病人的上界
$UB_j^e$	第 $j$ 个预约段结束时剩余病人的上界

根据先到先服务 (FCFS) 策略, 可以分别计算各个时段到达病人的期望等待费用:

$$W(s) = \sum_{m=0}^{a_1} \sum_{p=0}^{m-1} \beta(p) \partial_1(m) + \sum_{j=1}^{N-1} \sum_{x=0}^{UB_j^r} \sum_{m=0}^{a_{j+1}} \sum_{p=0}^{m-1} \beta(x+p) \theta_{j+1}(x, m) \quad (3)$$

其中  $\beta(\cdot)$  为每个病人关于等待时间的等待费用函数. (3) 式中第一项为第一个预约段等待时间产生的费用, 后面一项为第二段到第  $N$  段到达病人分别产生的期望等待费用.

文献中一般都假设  $\beta(t)$  为线性函数或者二次函数 [14,17,20]. 本文的模型是基于剩余人数概率分布直接计算每段到达病人所产生的期望等待费用, 对等待费用函数没有线性及凸性要求, 因此该模型适用于  $\beta(t)$  为任意函数的情形. 当其为线性函数时, 即  $\beta(t) = \omega t$ , 等待时间费用可进一步表示为:

$$W(s) = \sum_{m=0}^{a_1} \omega \frac{m(m-1)}{2} \partial_1(m) + \sum_{j=1}^{N-1} \sum_{x=0}^{UB_j^r} \sum_{m=0}^{a_{j+1}} \omega \left( mx + \frac{m(m-1)}{2} \right) \gamma_j(x) \partial_{j+1}(m) \quad (4)$$

各阶段剩余概率  $\gamma_j(x)$  的关联约束为:

$$\gamma_j(x) = \sum_{m=0}^{x+1} \partial_j(m) \gamma_{j-1}(x-m+1), \quad j = 1, 2, \dots, N, x \geq 1 \quad (5)$$

$$\gamma_j(0) = \partial_j(0) \gamma_{j-1}(0) + \sum_{m=0}^1 \partial_j(m) \gamma_{j-1}(-m+1), \quad j = 1, 2, \dots, N \quad (6)$$

$$\sum_{x=0}^{\max} \gamma_j(x) = 1, \quad j = 1, 2, \dots, N \quad (7)$$

其中为了表达统一, 令  $\gamma_0(0) = 1, UB_0^r = 0$ .

### 2.2.3 系统超时的期望费用

由于该模型采用过度预约的策略来弥补爽约, 因此系统可能会出现超时运行情况. 根据表 1 的参数定义, 系统的超时期望费用为:

$$O(s) = \sum_{x=0}^{UB_N^r} \varphi(x) \gamma_N(x) \quad (8)$$

其中  $\varphi(\cdot)$  为系统关于超时人数的超时费用函数, 与等待费用类似, 本文的模型同样适用于  $\varphi(t)$  为任意函数的情形, 当其为线性函数时, 即  $\varphi(x) = \tau x$ , 超时费用可进一步表示为:

$$O(s) = \sum_{x=0}^{UB_N^r} \tau x \gamma_N(x) \quad (9)$$

综上所述, 当系统等待费用和超时费用均为线性函数时, 系统优化模型为:

$$\begin{aligned} f^* &= \max B(s) - W(s) - O(s) \\ \text{s.t.:} & (5) \sim (7) \end{aligned} \quad (10)$$

其中  $B(s)$ 、 $W(s)$  及  $O(s)$  的计算表达式分别为 (2)、(4) 及 (9).

## 3 拉格朗日松弛算法

在前述模型中, 当  $S$  给定时,  $\gamma_j(x)$  的取值由约束 (5)~(7) 唯一确定. 为简化表达, 可以将约束 (5)~(6) 写成如下的一般形式:

$$\gamma_j(x) = \sum_{m=0}^{UB_{j-1}^r} g_{j-1}(x, m) \gamma_{j-1}(m) \quad (11)$$

即第  $j$  个时段剩余人数概率为第  $j-1$  时段剩余人数概率向量的线性组合, 组合系数与  $a_j$  的取值及  $x$ ,  $m$  有关, 具体计算公式如下.

$$g_{j-1}(x, m) = \begin{cases} \partial_j(x-m+1) & j \geq 1, UB_j^r \geq x \geq 1, 0 \leq m \leq UB_{j-1}^r \\ \partial_j(0) + \partial_j(1) & j \geq 1, x = 0, m = 0 \\ \partial_j(0) & j \geq 1, x = 0, m = 1 \leq UB_{j-1}^r \\ 0 & j \geq 1, x = 0, 2 \leq m \leq UB_{j-1}^r \end{cases} \quad (12)$$

重新将上面的优化问题表示为:

$$\begin{aligned} f^* &= \max f(s) \\ \text{s.t.: } \gamma_j(x) &= \sum_{m=0}^{UB_j^r} g_{j-1}(x, m) \gamma_{j-1}(m) \quad j = 1, 2, \dots, N, x = 0, \dots, UB_j^r \\ \sum_{x=0}^{UB_j^r} \gamma_j(x) &= 1 \quad j = 1, 2, \dots, N \end{aligned} \quad (13)$$

该优化问题的求解难度主要在于约束 (12), 因此本文提出了基于拉格朗日松弛的方法. 给定拉格朗日乘子  $u_{jx}$ , 松弛约束 (12), 对应的拉格朗日松弛函数为:

$$\begin{aligned} L(s, u) &= \sum_{j=1}^N \pi(1 - \sigma) a_j - \sum_{m=0}^{a_1} \omega \frac{m(m-1)}{2} \partial_1(m) - \sum_{j=1}^{N-1} \sum_{x=0}^{UB_j^r} \sum_{m=0}^{a_{j+1}} \omega \left( mx + \frac{m(m-1)}{2} \right) \gamma_j(x) \partial_{j+1}(m) \\ &\quad - \sum_{x=0}^{UB_N^r} \tau \gamma_N(x) x + \sum_{j=1}^N \sum_{x=0}^{UB_j^r} u_{jx} \left[ \gamma_j(x) - \sum_{m=0}^{UB_{j-1}^r} g_{j-1}(x, m) \gamma_{j-1}(m) \right] \end{aligned} \quad (14)$$

对应的拉格朗日松弛问题为:

$$\begin{aligned} L^*(u) &= \max_s L(s, u) \\ \sum_{x=0}^{UB_j^r} \gamma_j(x) &= 1 \quad j = 1, 2, \dots, N \end{aligned} \quad (15)$$

当  $s$  和  $u$  给定时, 松弛函数  $L(s, u)$  可以看成是  $\gamma_j(x)$  的线性组合, 组合系数与给定的  $s$  和  $u$  有关, 为表述方便, 可以重写松弛函数:

$$L(s, u) = \sum_{j=1}^N h(a_j) + \sum_{j=1}^N \sum_{x=0}^{UB_j^r} k_j(x) \gamma_j(x) \quad (16)$$

其中

$$h(a_j) = \begin{cases} \pi(1 - \sigma) a_1 - \sum_{m=0}^{a_1} \tau \frac{m(m-1)}{2} \partial_1(m) - \sum_{x=0}^{UB_1^r} u_{1x} g_0(x, 0), & j = 1 \\ \pi(1 - \sigma) a_j, & j > 1 \end{cases} \quad (17)$$

$$k_j(x) = \begin{cases} u_{jx} - \sum_{m=0}^{a_{j+1}} \tau \left( mx + \frac{m(m-1)}{2} \right) \partial_{j+1}(m) - \sum_{m=0}^{UB_{j+1}^r} u_{(j+1)m} g_j(m, x), & 1 \leq j \leq N-1 \\ u_{Nx} - \tau x, & j = N \end{cases} \quad (18)$$

给定  $u$ ,  $\gamma_j(x)$  的系数只与  $a_{j+1}$  有关, 因此该松弛问题可以用动态规划求解. 将  $a_j$  作为状态, 考虑如下的动态规划函数:

$$L_i(y, u) = \max_{\substack{a_{i+1}, \dots, a_N \\ \gamma_i, \dots, \gamma_N}} \left\{ \sum_{j=i}^N h(a_j) + \sum_{j=i}^N \sum_{x=0}^{UB_j^r} k_j(x) \gamma_j(x) \mid \sum_{x=0}^{UB_j^r} \gamma_j(x) = 1, \quad j = i, \dots, N, a_i = y \right\} \quad (19)$$

则松弛问题最优值与动态规划函数有如下关系:

$$L^*(u) = \max_{0 \leq y \leq UB_1^r} L_1(y, u) \quad (20)$$

不难得出上述动态规划函数有如下的递推关系:

$$\begin{aligned} &L_i(y, u) \\ &= \max_{0 \leq z \leq UB_{i+1}^r} \left\{ \max_{\substack{a_{i+2}, \dots, a_N \\ \gamma_i, \dots, \gamma_N}} \left\{ \sum_{j=i}^N h(a_j) + \sum_{j=i}^N \sum_{x=0}^{UB_j^r} k_j(x) \gamma_j(x) \mid \sum_{x=0}^{UB_j^r} \gamma_j(x) = 1, \quad j = i, \dots, N, a_i = y, a_{i+1} = z \right\} \right\} \\ &= \max_{0 \leq z \leq UB_{i+1}^r} \left\{ \max_{\substack{a_{i+2}, \dots, a_N \\ \gamma_{i+1}, \dots, \gamma_N}} \left\{ \sum_{j=i+1}^N h(a_j) + \sum_{j=i+1}^N \sum_{x=0}^{UB_j^r} k_j(x) \gamma_j(x) \mid \sum_{x=0}^{UB_j^r} \gamma_j(x) = 1, \quad j = i+1, \dots, N, a_{i+1} = z \right\} \right\} \end{aligned}$$

$$\begin{aligned}
& +h(a_i = y) + \max_{\gamma_i} \left\{ \sum_{x=0}^{UB_i^r} k_i(x)\gamma_i(x) \left| \sum_{x=0}^{UB_i^r} \gamma_i(x) = 1, a_{i+1} = z \right. \right\} \\
& = \max_{0 \leq z \leq UB_{i+1}^s} \left\{ L_{i+1}(z, u) + h(a_i = y) + \max_{0 \leq x \leq UB_i^r} k_i(x) \right\} \\
& = \max_{0 \leq z \leq UB_{i+1}^s} \left\{ L_{i+1}(z, u) + \max_{0 \leq x \leq UB_i^r} k_i(x) \right\} + h(a_i = y)
\end{aligned} \tag{21}$$

且

$$L_N(y, u) = \max_{\gamma_N} \left\{ h(a_N) + \sum_{x=0}^{UB_N^r} k_N(x)\gamma_N(x) \left| \sum_{x=0}^{UB_N^r} \gamma_N(x) = 1, a_N = y \right. \right\} = h(a_N) + \max_{0 \leq x \leq UB_N^r} \{u_{Nx} - \tau x\} \tag{22}$$

拉格朗日对偶问题为:

$$D^* = \min_u L^*(u) \tag{23}$$

该对偶问题的解构成了原问题的一个上界. 该对偶问题的求解可以采用经典的次梯度算法<sup>[21]</sup>.

定义算法相关参数及变量如表 2. 则拉格朗日松弛算法步骤如下:

**Step 1** 初始化算法的输入参数, 包括  $N, \pi, \sigma, \tau, \omega, \varepsilon$  及最大迭代次数  $\max\_iter$ ; 初始化每段预约人数上界  $UB_j^s$  及剩余生人数的上界  $UB_j^r$ ; 令  $k \leftarrow 1$ , 初始化拉格朗日乘子  $u$ .

**Step 2** 调用动态规划子程序求解松弛问题的目标函数值  $curObj$ .

**Step 3** 如果  $curObj < bestDualObj$ , 更新最优目标函数值  $bestDualObj \leftarrow curObj$ ; 否则转 Step 6;

**Step 4** 松弛解的可行化处理. 计算每段的剩余人数概率  $r_j(x)$ , 求取原问题目标函数值  $bestOriObj$ , 并更新原问题最优目标值  $bestOriObj$  及最优松弛调度序列  $bestRelaxScheduling$ .

**Step 5** 如果  $bestDualObj - bestOriObj < \varepsilon$ , 则最优近似解找到, 转 Step 8;

**Step 6** 计算次梯度  $\xi$ , 更新拉格朗日乘子  $u$ .

**Step 7** 如果  $k \geq \max\_iter$ , 转 Step 8; 否则  $k \leftarrow k + 1$ , 转 Step 2;

**Step 8** 输出最优对偶值  $bestDualObj, bestOriObj, bestRelaxScheduling$ .

表 2 拉格朗日松弛算法相关参数及变量

变量	定义
$bestDualObj$	最优对偶目标值
$bestOriObj$	最优可行解目标值
$curObj$	当前对偶目标值
$bestRelaxScheduling$	最优松弛调度序列
$\max\_iter$	最大迭代次数
$\varepsilon$	迭代终止的对偶间隙阈值
$k$	迭代次数
$u$	拉格朗日乘子
$\xi$	次梯度

从上述算法步骤可以看到该算法的计算量集中在动态规划算法及原问题目标值的计算. 对于动态规划算法, 其状态矩阵规模为  $O(N \cdot UB^s)$ , 其中  $UB^s$  为每段预约人数的上界, 不难发现, 一个保守的上界可以取为  $\lceil \frac{2\pi}{\omega(1-\sigma)} + 1 \rceil$ , 即与问题规模  $N$  无关, 每一个状态的计算是在  $UB^s + 1$  个量中取最大, 计算量为  $O(1)$ , 因此动态规划算法的复杂度为  $O(N)$ . 对于原问题目标函数值的计算, 公式 (2) 关于服务病人收益的计算复杂性为  $O(N)$ ; 公式 (3) 关于等待费用的计算是分别对各段剩余人数和到达人数进行期望求和, 因此计算量为  $O(N \cdot UB^r \cdot UB^s)$ , 其中  $UB^r$  为每段剩余人数的上界, 一个保守的上界为  $N \cdot UB^s$ , 根据前面关于  $UB^s$  的分析, 可知总的等待费用计算量为  $O(N^2)$ ; 公式 (8) 关于超时费用的计算量为  $O(N)$ , 这样可以得出系统总效用计算的复杂性为  $O(N) + O(N^2) + O(N) = O(N^2)$ , 因此整个算法的计算复杂度为  $O(N) + O(N^2) = O(N^2)$ .

#### 4 实验论证及分析

拉格朗日松弛算法的实验参数设置如下: 预约段数  $N$  取值为 4, 6, 8, 10, 12, 14, 16, 18, 爽约率为 0.2, 0.4, 0.6, 0.8, 算法的最大迭代次数为 500. 为了验证预约调度系统中不同参数对算法性能的影响, 分别设计



了 9 种不同的系统收益及费用调节因子组合, 服务病人收益因子、单位等待费用因子、系统超时费用因子 ( $\pi, \omega, \tau$ ) 取值分别为 (1, 1, 1)、(2, 1, 1)、(4, 1, 1)、(1, 2, 1)、(1, 4, 1)、(1, 1, 2)、(1, 1, 4)、(1, 1, 6)、(1, 6, 1), 一共设计 864 组实验组合.

表 3 拉格朗日松弛算法相关参数及变量

预约段数	爽约率	预约上界	预约调度	对偶目标值	原始问题值
4	0.2	3	1 1 1 1	3.200	3.200
	0.4	4	2 1 1 2	2.458	2.400
	0.6	4	2 2 2 2	2.145	1.957
	0.8	4	4 4 4 3	1.766	1.560
6	0.2	3	1 1 1 1 1 1	4.800	4.800
	0.4	4	2 0 2 2 2 1	3.693	3.600
	0.6	6	2 2 2 2 2 2	3.427	3.027
	0.8	6	4 4 4 4 4 4	2.695	2.329
8	0.2	3	1 1 1 1 1 1 1 1	6.400	6.400
	0.4	4	2 0 2 2 0 2 2 1	6.093	5.422
	0.6	6	2 2 2 2 2 2 2 2	4.509	3.982
	0.8	6	4 0 4 4 4 4 4 3	3.761	3.25
10	0.2	3	1 1 1 1 1 1 1 1 1 1	8.000	8.000
	0.4	4	2 2 2 0 2 2 2 2 0 2	6.164	6.000
	0.6	6	2 2 2 0 2 2 2 2 2 2	5.255	4.782
	0.8	10	4 0 4 4 4 4 4 4 4 3	4.552	3.910

通常拉格朗日松弛算法解的有效性用对偶间隙来描述. 表 3 给出了在基准组合 (1, 1, 1) 下算法的具体求解结果 (为表达简洁, 只列出预约段数  $N \leq 10$  的情形). 由于整数约束的存在, 拉格朗日松弛算法找到对偶问题最优值和原问题的最优值之间可能存在固有对偶间隙. 图 1 及图 2 分别给出了在不同参数组合下对偶间隙随爽约率及等待费用因子的变化情况, 从图可以观察到如下现象:

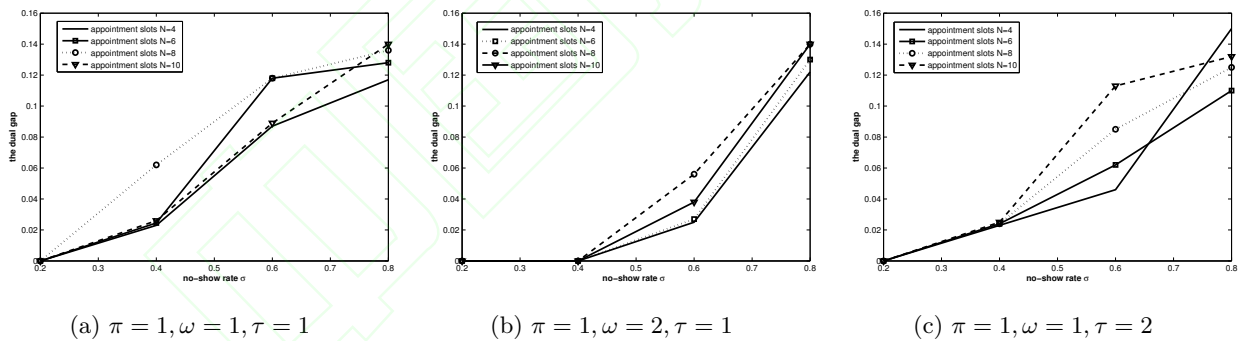


图 1 对偶间隙随爽约率的变化趋势

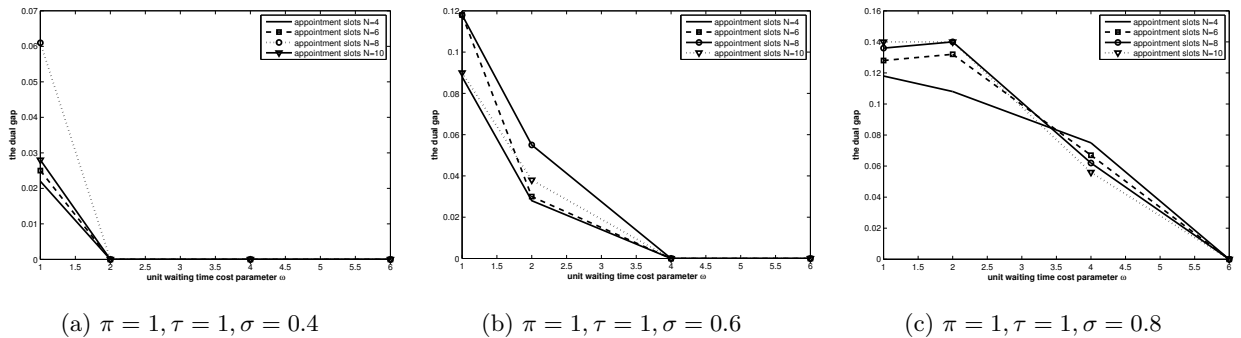


图 2 对偶间隙随等待因子的变化趋势

1、在预约调度的研究中, 爽约率是影响系统性能的重要因素. 从图 1 可以看出, 当爽约率较低时, 对偶间隙较低, 这意味着本文提出的算法都能找到很好的近似解, 统计结果显示, 在 864 组实验组合中, 当爽约率为 0.2 时算法都找到了最优解. 随着爽约率的增大, 对偶间隙增大, 意味算法找到的近似解质量变差, 但基本控制在 15% 以内, 同时由于整数约束造成的固有对偶间隙, 近似值与原问题最优值之间的真实间隙实际都小

于对偶间隙.

2、在设计实际的预约系统时, 病人的等待时间是一个重要的参考量, 等待时间过长, 病人的满意度下降. 当医疗服务提供者更加看重病人等待影响时, 可以加大等待因子的比重. 从图 2 可以看出, 随着单位等待费用的增加, 对偶间隙降低, 即算法找到了更好的近似解. 原因在于单位等待费用增加, 每个预约时段可预约人数的上界减小, 也即可行解的空间缩小了.

LaGanga<sup>[14]</sup> 中运用启发式算法寻优最优目标函数, 并指出算法求解质量依赖于初始化调度. 本文进一步设计了松弛 - 启发式算法, 首先通过拉格朗日松弛算法得到原问题可行解, 然后以此为初始化序列, 运用文献 [14] 中启发式算法进行邻域搜索. 在随机产生的 864 个随机实例中, 当爽约率为 0.2 时, 两种算法都能找到最优解, 统计结果显示, 纯启发式算法优于松弛 - 启发式算法与松弛 - 启发式算法优于纯启发式算法比为 1:3. 图 3 进一步在  $(\pi, \omega, \tau) = (2, 1, 4)$  的参数组合下, 比较了两种算法在不同爽约率下系统总效用目标值随预约时段的变化趋势. 可以看出, 在不同的爽约率下, 松弛 - 启发式算法找到的最好解整体上优于纯启发式算法, 尤其是在较大的爽约率下, 这种优势更为明显, 这也说明拉格朗日松弛算法找到的可行解在解空间上离最优解相对较近, 提高了启发式算法的搜索能力.

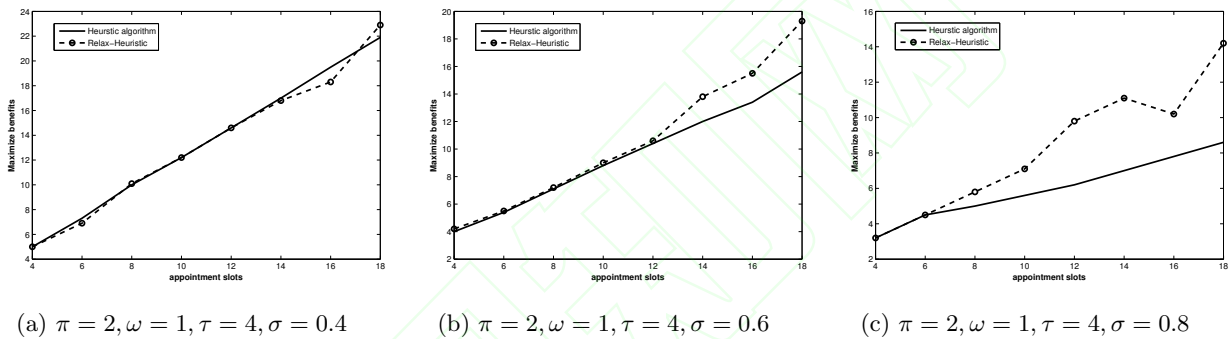


图 3 纯启发式算法与松弛 - 启发式算法求解质量对比

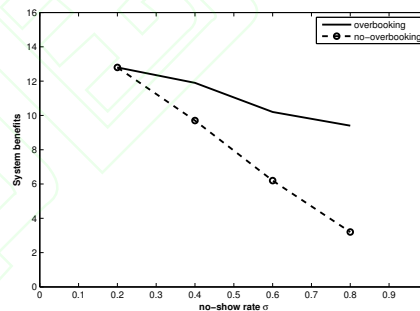


图 4 过度预约策略与非过度预约策略下系统的总效用随爽约率的变化趋势

本文进一步研究了过度预约 (overbooking) 策略对预约系统总效用的影响. 当不采用过度预约策略时, 即每个时段只预约一个病人时, 这时没有等待和超时出现, 易得系统的总效用为  $\pi N(1 - \sigma)$ , 即其关于爽约率  $\sigma$  是线性降低的. 当采用过度预约时, 图 4 给出了系统总效用的变化趋势, 随着爽约率的增加, 系统效用呈下降趋势, 但下降幅度明显小于没用采用过度预约的情况. 这说明了系统的过度预约策略能在一定的程度上降低爽约给系统性能带来的不利影响.

## 5 结论

在现实的预约系统中, 爽约是影响预约系统的重要因素. 因此, 如何在考虑预约患者存在爽约的情形下, 设计合理有效的调度策略以及优化算法来保证医院的收益, 是一个值得研究的问题. 本文针对医疗门诊等服务行业预约系统的爽约问题, 构建了基于过度预约的优化模型, 模型以预约调度人数为优化变量, 由于每个预约时段剩余人数与下一个时段实际到达人数之间的独立性, 每段开始时等待人数的概率分布可以直接计算, 进而可以直接计算每段到达人数的期望等待费用, 对等待费用函数没有线性及凸性要求. 为了克服剩余概率耦合约束带来的计算困难, 本文提出了拉格朗日松弛求解算法, 设计了动态规划算法求解相应的松弛问题, 采

用经典次梯度法求解对偶问题. 数值实验表明, 随着爽约率的增加, 系统效用会降低. 针对小规模的预约段数时, 本文提出的算法基本都能找到最优解. 在预约段数为中等规模时, 算法找到近似解整体优于文献中提出的算法. 同时实验还进一步分析了爽约率对系统性能的影响, 指出过度预约策略能在一定的程度上降低爽约给系统性能带来的不利影响. 在后续的研究中, 可以考虑将该模型拓展到含有临时到达需求的预约调度问题, 同时还可以考虑多服务资源的情形.

## 参考文献

- [1] Zacharias C, Pinedo M. Appointment scheduling with no-shows and overbooking[J]. *Production and Operations Management*, 2014, 23(5): 788–801.
- [2] Ho C J, Lau H S. Minimizing total cost in scheduling outpatient appointments[J]. *Management Science*, 1992, 38(12): 1750–1764.
- [3] Begen M A, Queyranne M. Appointment scheduling with discrete random durations[J]. *Mathematics of Operations Research*, 2011, 36(2): 240–257.
- [4] Tang J F, Yan C J, Cao P P. Appointment scheduling algorithm considering routine and urgent patients[J]. *Expert Systems with Applications*, 2014, 41(10): 4529–4541.
- [5] 杜少甫, 谢金贵, 刘作仪. 医疗运作管理: 新兴研究热点及其进展 [J]. *管理科学学报*, 2013, 16(8): 1–19  
Du Shaofu, Xie Jingui, Liu Zuoyi. Progress and prospects in an emerging hot topic: Healthcare operations management[J]. *Journal of Management Sciences in China*, 2013, 16(8): 1–19.
- [6] Bailey N T, Welch J D. Appointment systems in hospital outpatient departments[J]. *Operational Research Society*, 1952, 15(3): 224–232.
- [7] Cayirli T, Veral E. Outpatient scheduling in healthcare: A review of literature[J]. *Production and Operations Management*, 2003, 12(4): 519–549.
- [8] Gupta D, Denton B. Appointment scheduling in health care: Challenges and opportunities[J]. *IIE Transactions*, 2008, 40(9): 800–819.
- [9] Bhattacharjee P, Ray P K. Modelling and optimisation of healthcare systems: An overview[J]. *International Journal of Logistics Systems and Management*, 2014, 19(3): 356–371.
- [10] 罗利, 石应康. 医疗服务资源调度优化理论、方法及应用 [M]. 北京: 科学出版社, 2014.  
Luo Li, Shi Yingkang. Optimization healthcare resource planning and scheduling modeling: Theory, methods and applications[M]. Beijing: Science Press, 2014.
- [11] Rust C T, Gallups N H, Clark W S, et al. Patient appointment failures in pediatric resident continuity clinics[J]. *Archives of Pediatrics & Adolescent Medicine*, 1995, 149(6): 693–695.
- [12] Robinson L W, Chen R R. A comparison of traditional and open-access policies for appointment scheduling[J]. *Manufacturing & Service Operations Management*, 2010, 12(2): 330–346.
- [13] LaGanga L, Lawrence S R. Yield management in health care clinics with walk-in traffic[C]// *Proceedings of Tradition and Innovation in Operations Management, 15th International Annual EurOMA Conference*, University of Groningen, the Netherlands, 2008.
- [14] LaGanga L R, Lawrence S R. Appointment overbooking in health care clinics to improve patient service and clinic performance[J]. *Production and Operations Management*, 2012, 21(5): 874–888.
- [15] Muthuraman K, Lawley M. A stochastic overbooking model for outpatient clinical scheduling with no-shows[J]. *IIE Transactions*, 2008, 40(9): 820–837.
- [16] 曹萍萍, 唐加福. 考虑存在取消预约情形的门诊预约能力分配策略 [J]. *运筹与管理*, 2014, 23(2): 250–257.  
Cao Pingping, Tang Jiafu. Capacity allocation policy of outpatient appointment considering cancellation of appointment[J]. *Operations Research and Management Science*, 2014, 23(2): 250–257.
- [17] Denton B, Gupta D. A sequential bounding approach for optimal appointment scheduling[J]. *IIE Transactions*, 2003, 35(11): 1003–1016.
- [18] Kaandorp G C, Koole G. Optimal outpatient appointment scheduling[J]. *Health Care Management Science*, 2007, 10(3): 217–229.
- [19] Kim S, Giachetti R E. A stochastic mathematical appointment overbooking model for healthcare providers to improve profits[J]. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 2006, 36(6): 1211–1219.
- [20] 阎崇钧, 唐加福, 姜博文. 带临时需求的预约系统最优能力计划与调度策略 [J]. *东北大学学报*, 2014, 35(8): 1083–1087.  
Yan Congjun, Tang Jiafu, Jiang Bowen. Optimal capacity and scheduling policy for appointment system with walk in demand[J]. *Journal of Northeastern University (Natural Science)*, 2014, 35(8): 1083–1087.
- [21] Kim S, Ahn H, Cho S C. Variable target value subgradient method[J]. *Mathematical Programming*, 1990, 49(1–3): 359–369.