



# 带有响应时间承诺的 选址 - 分配问题研究

胡丹丹<sup>1</sup>, 杨超<sup>1</sup>, 刘智伟<sup>2</sup>

1 华中科技大学 管理学院, 武汉 430074

2 华中科技大学 控制科学与工程系, 武汉 430074

**摘要:** 研究在服务行业中一类带有响应时间承诺的设施选址及对各选址点进行服务台数目分配的问题。将排队理论与选址理论有机结合, 建立带有预算限制的非线性整数规划模型, 以找到合适的选址地点和相应的服务台数目, 使在承诺的响应时间内完成服务的需求数目最大。运用贪婪算法得出服务台分配的结果, 采用下降算法和基于二进制编码的遗传算法对设施点选址, 通过大量随机算例验证算法的有效性。数值模拟结果表明, 目标值随着预算或承诺时间的增加而增加, 但是随着需求数量的增加不是呈单调递增趋势, 而且随着一些参数的变化平均每个需求的平均逗留时间和平均运输时间的波动基本上是相反的。

**关键词:** 选址; 排队; 多服务台; 启发式算法

**中图分类号:** C93

**文献标识码:** A

**文章编号:** 1672-0334(2010)01-0114-08

## 1 引言

传统的选址问题包括中位问题、覆盖问题和中心问题, 这几类选址问题只关注顾客与设施之间的距离, 如覆盖半径、运输费用等<sup>[1]</sup>。随着竞争日益激烈, 人们生活工作的步伐加快, 时间已经成为成败的重要因素, 人们对时间的关注程度越来越高。顾客光临某设施时, 所花费的时间不仅仅是在路上的行进时间, 还包括由于设施拥塞所造成的等待时间, 因此顾客对设施的选择不仅考虑到与设施之间的距离, 还考虑到设施的拥塞程度和等待时间等因素。近年来, 将排队等待时间、系统逗留时间等排队问题引入拥塞设施选址 - 分配问题的研究在国际上已引起了广泛关注<sup>[2]</sup>。

## 2 相关研究评述

较早的排队选址问题主要是研究如何对一个设施进行选址<sup>[3-5]</sup>, 近年来对排队选址问题的研究主要考虑对多个设施进行选址, 并提供相应的启发式算法求解<sup>[2]</sup>。研究排队选址问题主要有两类模型, 即

排队覆盖模型和排队中位模型。排队覆盖模型考虑的是如何提供足够的覆盖空间, 排队中位模型考虑的是如何减少交通费用。

排队中位模型的研究已经比较广泛。Silva 等将 M/M/1 排队系统应用于连锁行业竞争选址研究, 将等待时间引入到竞争模型中, 目标函数是使交通费用和等待时间最小<sup>[6]</sup>; Wang 等考虑的模型与 Silva 相似, 假设顾客选择最近的设施服务, 并对最大期望等待时间加以约束, 对单服务台的一个设施进行选址使交通费用和等待时间最小<sup>[7]</sup>; Zhang 等研究竞争设施的选址问题, 将顾客等待时间作为约束条件, 需求分配与设施容量成正比、与距离成反比, 通过选址和容量调整, 使顾客的空间效用最大, 其中设施容量是离散的, 与服务台数目相关<sup>[8]</sup>; Marianov 等研究多服务台的设施在竞争环境下的选址, 需求分配与顾客到达设施所用时间和等待服务时间之和成反比, 并结合禁忌搜索和贪婪随机适应搜索算法求解<sup>[9]</sup>。排队中位模型还广泛应用于物流、供应链上的设施选址问题。Benjaafar 等研究带库存的 M/M/1 排队生产

收稿日期: 2009-02-24 修返日期: 2009-10-26

基金项目: 国家自然科学基金(70601011, 70871044); 教育部新世纪优秀人才支持计划项目(NCET-06-0653)

作者简介: 胡丹丹(1983-), 女, 浙江慈溪人, 华中科技大学管理学院博士研究生, 研究方向: 网络与决策优化等。

E-mail: hudandan1224@163.com

系统,不仅要考虑如何进行需求分配,还要决定最优库存水平,使运输费用、库存费用和缺货费用之和达到最小,并将问题拓展到带有固定选址费用的选址模型<sup>[10]</sup>;Huang等研究类似分配中心、物流中转站之类的设施选址问题,将该设施模拟为M/G/1排队系统,设施容量由各个系统的服务时间期望值和方差决定,考虑设施容量固定和可变两种情形下选址-分配问题<sup>[11]</sup>;Lieckens等将排队论应用于逆向物流选址问题<sup>[12]</sup>。此外,还有一类排队选址问题主要研究超立方排队模型<sup>[13,14]</sup>。

排队覆盖模型的研究基本上沿着Marianov等<sup>[15]</sup>的方向进行。假设各个设施是M/M/c排队系统,潜在顾客流失情况分为两类,一类是由于需求点没有在设施的覆盖半径之内而流失的顾客,一类是由于拥塞而导致顾客减少,模型目标是使设施所能服务的平均顾客率最大。在此基础上,Shavandi等应用模糊理论,研究M/M/1排队系统选址问题<sup>[16]</sup>;Berman等考虑在流失顾客不超过某预定值的条件下,使建立的设施数目最少问题<sup>[17]</sup>。目前,对于排队覆盖选址问题的研究较少,也相对不够成熟和全面,本研究将以排队覆盖模型为基础进行进一步研究。

在现实中,众多商家和企业为了增强竞争力向顾客承诺响应时间,如外卖店,除了食品质量,顾客最为关注的就是响应时间。永和大王将响应时间降低到20分钟以内,为其赢得了良好声誉;米诺比萨外卖给出在30分钟内送货上门,否则顾客免费享用的承诺;也有很多外卖店承诺在响应时间之外给予打折优惠。此外,在供应链上由订单驱动的制造企业也受响应时间的约束。商家为了保证服务质量和良好信誉,需尽量在承诺的响应时间完成服务,其中响应时间包括订单在系统中的逗留时间和产品运输时间。对于这一类设施的选址研究具有重要的现实意义,目前基本上没有文献对这类设施选址问题进行研究。

本研究建立非线性整数规划模型,由于选址问题是NP困难问题,不可能在多项式时间内计算完成,对于大规模问题不能用精确算法求解,本研究提出启发式算法求解。本研究的模型目标是在某区域内找到最优选址地点和相应的服务台数目(服务效率)使服务的顾客数目最大,顾客是指在设施覆盖范围之内并且能在最大响应时间之内得到服务的需求。模型决策变量包括选址决策和服务台分配决策,服务台这个概念比较广泛,可以指服务人员、生产设备等,服务台数目代表服务效率的高低。目前,将服务台数目做为决策变量的文献研究非常少<sup>[18,19]</sup>。由于资金有限,若选取多个设施选址点,每个设施设置的服务台数目就较少,虽然可以缩短运输时间,但会增加订单在系统中的排队等待时间,如果减少设施数目,单个设施内的服务台数目增加,减少排队等待时间,但是运输时间就会加长。本研究通过启发式算法给出选址和设置服务台这两个决策的较优结果,并通过随机算法验证启发式算法的有效性。

### 3 模型建立

根据本研究需要给出如下假设。①需求点和候选设施选址点都是位于平面上的离散点;②各个服务设施之间是互相独立的,不存在相互合作的关系;③所有设施提供的商品或服务是同质的,所以服务台的工作效率是相同的;④一个服务台被一个订单占用,若有其他订单到达且没有空闲服务台时,列入等待队列;⑤各点的需求率(单位时间内的需求量)服从泊松分布,各个服务台的服务时间服从负指数分布,采用先来先服务原则;⑥顾客事先不知道各个设施的繁忙情况,所以都选择最近的设施为其服务;⑦考虑预算限制。预算费用包括选址费用和服务台费用,选址费用是指建设费用或租赁费用等,不同的选址点选址费用不同;服务台费用与服务台数目成正比,且不同选址点设施内的单个服务台费用相同。

本研究定义的集合、参数和决策变量列示如下。

#### (1) 集合

$I$  为需求点集合,  $i \in I$ ;

$J$  为候选设施选址点集合,  $j \in J$ 。

#### (2) 参数

$f_i$  为点  $i$  上单位时间内产生的需求(需求率),服从参数为  $\lambda_i$  的泊松分布;

$\bar{\lambda}_j$  为位于选址点  $j$  的设施单位时间内处理的平均需求;

$w_j$  为订单在选址点  $j$  的设施内的逗留时间,包括排队等待时间和在服务台上的处理时间;

$t_{ij}$  为选址点  $j$  上的设施对需求点  $i$  的响应时间;

$T$  为承诺的最大响应时间;

$d_{ij}$  为点  $i$  与点  $j$  之间的运输时间;

$B$  为预算资金;

$a_j$  为在选址点  $j$  上建设设施的选址费用;

$b$  为单个服务台费用;

$\mu$  为单个服务台单位时间内处理的平均需求。

#### (3) 决策变量

$$y_j = \begin{cases} 1 & \text{将设施建在候选设施选址点 } j \text{ 上 } (j \in J) \\ 0 & \text{否则} \end{cases}$$

$$x_{ij} = \begin{cases} 1 & i \text{ 点的需求分配给位于选址点 } j \text{ 上的设施} \\ 0 & \text{否则} \end{cases}$$

$m_j$  为位于选址点  $j$  上的设施内的服务台数目;

$P(\cdot)$  为概率分布函数。

根据以上各符号和变量的定义,建立选址模型。

$$\max Z = E \left[ \sum_{i \in I} \sum_{j \in J} f_i P(t_{ij} \leq T) x_{ij} \right] \quad (1)$$

s. t.

$$x_{ij} \leq y_j \quad \forall i \in I, j \in J \quad (2)$$

$$\sum_{j \in J} x_{ij} = 1 \quad \forall i \in I \quad (3)$$

$$\sum_{k \in I, d_{ik} \leq d_{ij}} x_{ik} \geq y_j \quad \forall i \in I, j \in J \quad (4)$$

$$\sum_{j \in J} (a_j y_j + b m_j) \leq B \quad \forall i \in I \quad (5)$$

$$m_j \leq \delta y_j \quad \forall j \in J \quad (6)$$

$$t_{ij} = d_{ij} + w_j \quad \forall i \in I, j \in J \quad (7)$$

$$\bar{\lambda}_j = E(\sum_{i \in I} f_i x_{ij}) = \sum_{i \in I} \lambda_i x_{ij} \quad \forall j \in J \quad (8)$$

$$\bar{\lambda}_j < m_j \mu \quad \forall j \in J \quad (9)$$

$$x_{ij} \in \{0, 1\} \quad \forall i \in I, j \in J \quad (10)$$

$$y_j \in \{0, 1\} \quad \forall j \in J \quad (11)$$

$$m_j \in \mathbb{N} \quad \forall j \in J \quad (12)$$

目标函数(1)式表示满足响应时间不超过  $T$  的需求率的期望值最大;约束条件(2)式表示只有在选址点  $j$  建立设施后,才能将需求分配给该点上的设施;约束条件(3)式确保每一个需求点上的全部需求只能分配给一个设施;约束条件(4)式保证各需求点分配给最近的设施,  $k$  为符合约束条件  $d_{ik} < d_{ij}$ , 属于集合  $J$  中的元素;约束条件(5)式保证总费用不超过预算限制;约束条件(6)式表示只有建设设施后,才能分配给该设施服务台,  $\delta$  为一个很大的正数,可以取  $\delta = \lceil B/b \rceil$  ( $\lceil a \rceil$  表示大于等于  $a$  的最小整数,  $a$  是有理数);约束条件(7)式表示响应时间包括在路上的运输时间和订单在系统中的逗留时间;约束条件(8)式表示各个设施平均处理的需求率;约束条件(9)式保证系统稳定性,不会排成无限的长队;约束条件(10)式和约束条件(11)式是整数约束;约束条件(12)式保证服务台数目为自然数。

求解目标函数  $Z$  需要引入排队论相关理论。首先  $Z$  可以进行如下转化。

$$\begin{aligned} Z &= E[\sum_{i \in I} \sum_{j \in J} f_i P(t_{ij} \leq T) x_{ij}] \\ &= E[\sum_{i \in I} \sum_{j \in J} f_i P(w_j \leq T - d_{ij}) x_{ij}]^{T_{ij} = T - d_{ij}} \\ &= \sum_{i \in I} \sum_{j \in J} \lambda_i F(T_{ij}) x_{ij} \end{aligned}$$

$F(T_{ij})$  是一个概率分布函数,表示订单在系统中逗留时间不大于  $T_{ij}$  的概率。以下介绍在排队理论中一般的  $M/M/c$  排队系统中的一些参数的表达方式。

系统中没有顾客的概率为

$$p_0 = \left[ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!(1-\rho_c)} \right]^{-1}$$

有  $c$  个顾客的概率为

$$p_c = \frac{\rho^c}{c!} p_0$$

其中,  $\rho = \frac{\lambda}{\mu}$ ,  $\rho_c = \frac{\rho}{c}$ , 由于  $p_0$  的表示比较繁杂, 故用  $\rho$  和  $\rho_c$  来代替,  $\rho$  和  $\rho_c$  没有现实意义;  $\lambda$  为平均需求产生率;  $c$  为服务台的数目;  $\mu$  的含义与以上模型中一样, 表示单个服务台处理的平均需求率。则订单在系统中的逗留时间不大于  $t$  的概率  $F(t, c)$  为<sup>[20]</sup>

当  $c = 1$  时,  $F(t, 1) = 1 - e^{-(\mu-\lambda)t}$ ;

当  $c > 1$  时,

$$F(t, c) = \begin{cases} 1 - (1 + \frac{\rho_c \mu t}{1 - \rho_c}) e^{-\mu t}, \rho = c - 1 \\ 1 - [1 + \frac{\rho_c}{(c-1-\rho)(1-\rho_c)}] e^{-\mu t} + \frac{\rho_c}{(c-1-\rho)(1-\rho_c)} e^{-\rho c(1-\rho_c)t}, \rho \neq c - 1 \end{cases}$$

排队等待时间  $t$  的分布函数为

$$W_q(t, c) = 1 - \frac{P_c}{1 - \rho_c} e^{-\mu(c-\rho)t}$$

需求不能立刻得到处理的延迟概率为  $DT(c) = 1$

$$- W_q(0, c) = \frac{P_c}{1 - \rho_c}, \text{ 令 } g(c) = DT(c) e^{-\mu(c-\rho)t}。$$

引理 1  $g(c)$  是凸函数

证明:  $g''(c) = DT''(c) e^{-\mu(c-\rho)t} + \mu^2 DT(c) e^{-\mu(c-\rho)t} - 2DT'(c) e^{-\mu(c-\rho)t}$ , 由于  $DT(c)$  是关于  $c$  的非递增凸函数<sup>[21]</sup>, 即  $DT'(c) \leq 0, DT''(c) \geq 0$ , 所以  $g''(c) \geq 0$ 。

定理 1 系统逗留时间分布函数  $F(t, c)$  是关于  $c$  的凹函数

证明: 由引理 1 可知,  $g(c)$  是凸函数, 则排队等待时间分布函数  $W_q(t, c) = 1 - g(c)$  是关于参数  $c$  的凹函数, 即  $\frac{\partial^2 W_q(t, c)}{\partial c^2} \leq 0$ 。由于系统逗留时间  $w = q + v$ ,  $q$

为排队等待时间,  $v$  为在服务台上的处理时间, 则

$$\begin{aligned} F(t, c) &= P\{w \leq t\} = P\{q + v \leq t\} \\ &= \int_0^t P\{q \leq t - t_v\} dP\{v \leq t_v\} \end{aligned}$$

由于处理时间  $v$  服从负指数分布, 则

$$\begin{aligned} F(t, c) &= \int_0^t W_q(t - t_v, c) d(1 - e^{-\mu t_v}) \\ &= \int_0^t W_q(t - t_v, c) \mu e^{-\mu t_v} dt_v \end{aligned}$$

$$\frac{\partial^2 F(t, c)}{\partial c^2} = \int_0^t \frac{\partial^2 W_q(t - t_v, c)}{\partial c^2} \mu e^{-\mu t_v} dt_v$$

$$\text{由于 } \frac{\partial^2 W_q(t - t_v, c)}{\partial c^2} \leq 0, \text{ 所以 } \frac{\partial^2 F(t, c)}{\partial c^2} \leq 0, F(t, c)$$

是关于参数  $c$  的凹函数。

所以  $F(t, c)$  具有以下两个性质。

- (1)  $F(t, c)$  是关于参数  $c$  的非递减函数;
- (2)  $F(t, c)$  是关于参数  $c$  的凹函数, 即  $F(t, c + 1) - F(t, c)$  是关于  $c$  的非递增函数。

基于性质(1)和性质(2), 在给定所有设施选址点的情况下, 通过贪婪算法能够求得最优服务台分配结果<sup>[18,19]</sup>。

### 4 算法

选址问题是个 NP 困难问题, 如果是小规模算例, 可以通过枚举法得到最优解, 对于大规模的问题需要给出启发式算法。本研究提出的算法包括两部分, 一是利用  $F(t, c)$  的性质对服务台进行分配, 二是用启发式算法找到较优的选址点。本研究利用贪婪算法对服务台进行分配, 利用下降算法和遗传算法寻得选址点, 具体算法如下。

#### 4.1 服务台分配精确算法

已知系统逗留时间分布函数性质(1)和性质(2), 本研究通过以下贪婪算法求得最优服务台分配最优解<sup>[19]</sup>。已知所选的地点为  $S, S \subseteq J$ 。步骤如下所示。

- (1) 计算每个所选地点  $j$  的平均需求率  $\bar{\lambda}_j (j \in S)$ ;
- (2) 对所选每个地点上的设施分配服务台,  $k_j$

$= \lceil \frac{\bar{\lambda}_j}{\mu} \rceil, k_j$  为设施  $j$  分配到的服务台数目;

(3) 如果  $\sum_{j \in S} (bk_j + a_j) > B$ , 无可行解, 结束; 如果  $B - b < \sum_{j \in S} (bk_j + a_j) \leq B, k_i$  为其分配结果, 结束;

(4) 如果  $\sum_{j \in S} (bk_j + a_j) \leq B - b$ , 对剩下的服务台进行分配, 计算对每个选址点上的设施增加一个服务台时目标值的增量, 即  $\Delta_j = \sum_{i \in A(j)} \lambda_i [F(t_i, k_j + 1) - F(t_i, k_j)]$ ,  $A(j)$  为分配给选址点  $j$  上的设施的需求点集合, 将服务台分配给  $\Delta_j$  最大的选址点, 并令  $k_j = k_j + 1$ , 重复步骤(4), 直到所有服务台分配完。

### 4.2 下降算法

下降算法 (descent algorithm, DA) 通过搜索给定的设施选址点集合  $S$  的邻域  $S'$  来求得满意解, 邻域  $S'$  定义为从  $S$  中去掉一个选址点, 增加一个候选选址点加入  $S$ , 已经选择的选址点与未选择的候选选址点交换。结合上文计算目标值, 具体步骤如下。

(1) 随机选取两个选址点, 构成初始设施地点集合  $S$ , 目标值为  $Z^*$ , 如果  $S$  为不可行解,  $Z^* = -\text{inf}$  (负无穷);

(2) 搜索  $S$  的所有邻域  $S', S'$  目标值表示为  $Z(S')$ 。所有邻域中计算得到最大目标值为  $Z_{\max}(S'_{\max})$ , 对应的邻域为  $S'_{\max}$ ;

(3) 如果  $Z_{\max} > Z^*$ , 令  $Z^* = Z_{\max}, S = S'_{\max}$ , 返回(2), 否则, 停止计算,  $Z^*$  为所求目标值,  $S$  为选址结果。

### 4.3 遗传算法

本研究设计遗传算法 (genetic algorithm, GA), 通过交叉、选择、变异操作识别出好的染色体, 求得选址点。染色体的适应值为目标值, 通过贪婪算法分配服务台求得。交叉、变异操作和求解一般问题时一样, 本研究不详细讲解。算法的主要技术如下。

(1) 编码

采用 0-1 编码, 编码位数为  $|J|, |J|$  表示候选设

施选址点的数目, 0 为该地点不建设施, 1 为该地点建设服务台。随机产生  $n$  条染色体, 该种群作为初始种群。

(2) 交叉、选择、变异

按  $P_{\text{cross}}$  交叉概率对  $n$  条染色体进行单点交叉, 形成子代染色体, 将父、子代染色体合并为  $2n$  个染色体, 并删掉有重复的染色体, 形成新的种群  $\omega$ , 选取目标值最大的前  $n$  条染色体, 并将到目前为止取得的最优值保存到另一个数组中。当种群  $\omega$  的染色体数目小于  $n$  时, 随机产生  $n - |\omega|$  条染色体添加到种群  $\omega$  中 ( $|\omega|$  为种群  $\omega$  的染色体数目)。随机选择一条染色体, 按  $P_{\text{mutation}}$  的变异率选择变异点进行变异, 通过以上步骤形成新的父代。

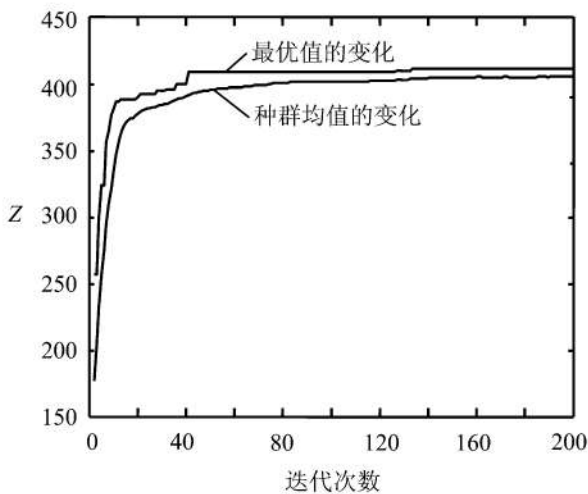
(3) 染色体可行性操作

无论是初始个体、交叉个体或是变异个体都有可能产生不可行解, 即不能满足预算限制约束条件(5)式。为了保持种群多样性, 避免陷入局部最优解中, 随机将编码为 1 的基因改为 0, 直到满足预算约束条件。

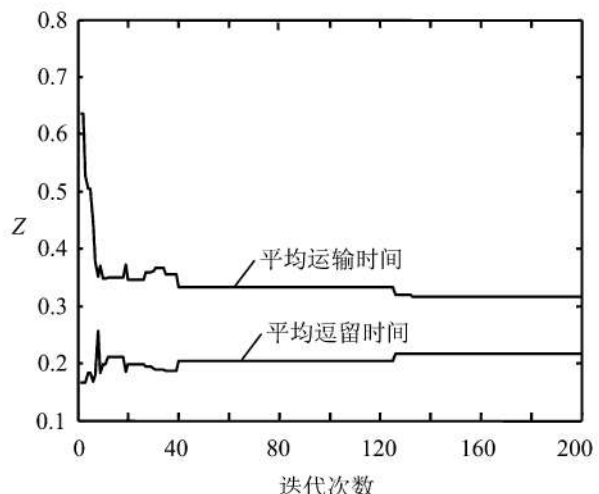
## 5 算例分析

本研究使用 Matlab 7.0 进行编程, 在 intel Pentium Dual E2200 2.2GHz、内存为 2.00GB 的计算机上进行计算试验。在  $[0, 5] \times [0, 5]$  的平面上随机产生需求点, 候选设施选址点为所有的需求点, 即  $J = I$ 。各点的选址费用为产生于区间  $[10, 20]$  的随机数, 各点的平均需求率服从  $[10, 20]$  均匀分布,  $\mu = 8, b = 8$ 。

遗传算法的参数设置为种群大小为 50, 交叉率为 0.8, 变异率为 0.03, 演化代数 200。其他参数设置为预算限制  $B = 900$ , 需求点数目  $|I| = 50$ , 承诺的最大响应时间  $T = 0.5$ 。图 1(a) 显示本研究设计的遗传算法的收敛情况令人满意, 图 1(b) 表示在迭代过程中平均每个需求在系统中的平均逗留时间和平均运输时间的变化状况, 纵坐标  $Z$  表示目标值。



(a)



(b)

图1 遗传算法进化过程

Figure 1 Evolutionary Process of Genetic Algorithm

为了得到启发式算法结果同最优值之间的差距,本研究通过枚举算法得到最优值。由于枚举算法运算时间是指数增长,所以必须选取小规模算例进行试验。 $T=0.5$ ,随机产生3组算例,每一组算例的点数与预算限制组合为 $\{|I|, B\} = \{20, 450\}, \{30,$

$650\}, \{40, 850\}$ ,并且 $\frac{B - \lceil \frac{\sum_{i \in I} \lambda_i}{\mu} \rceil b}{\min_{j \in J} a_j} \leq 6$ ,表示建立的设施点最多为6个。每组算例随机运行20次,其计算结果见表1。

由表1可知这两种算法的平均误差均不超过

4.4%,取得最优解的频率也较高。同时使用这两种算法得到较优解的目标值为 $Z_{better}, Z_{better}$ 基本上都是最优解。

表2比较了下降算法和遗传算法的计算结果。不同的需求点和预算限制组成一组算例,在每一组算例中,随机产生具有可行解的10个算例。由表2可知,下降算法和遗传算法计算效果都较好,其中遗传算法较容易找到可行解,而下降算法计算时间少。下面分析当 $B, |I|$ 和 $T$ 等参数变化时对结果的影响。由于采用启发式算法求解,求解结果会有一些的误差,仅对大致变化趋势进行观察分析。

表1 启发式算法与精确算法的比较

Table 1 Comparison between Heuristics and Exact Algorithm

点数	预算限制	下降算法		遗传算法		同时使用下降算法和遗传算法	
		平均误差	最优解	平均误差	最优解	平均误差	最优解
20	450	0.027	60%	0.000	100%	0.000	100%
30	650	0.012	90%	0.003	80%	0.000	100%
40	850	0.044	35%	0.014	80%	0.007	95%

注:同时使用下降算法和遗传算法得到的解是指由下降算法和遗传算法得到的两个结果中较优的那个解。

表2 下降算法和遗传算法计算结果

Table 2 Computational Results of Descent and Genetic Algorithms

点数	预算限制	下降算法					遗传算法				
		min	ave	max	$\frac{BK}{\text{Feasible}}$	Time(s)	min	ave	max	$\frac{BK}{\text{Feasible}}$	Time(s)
20	450	0.000	0.003	0.015	8/10	0.45	0.000	0.001	0.002	8/10	12.09
30	650	0.000	0.029	0.083	4/8	1.39	0.000	0.015	0.075	6/10	16.81
40	850	0.000	0.010	0.052	6/10	1.13	0.000	0.017	0.085	8/10	17.84
50	1 000	0.000	0.000	0.000	6/6	1.43	0.000	0.029	0.087	8/10	19.36
60	1 200	0.000	0.000	0.000	10/10	1.71	0.000	0.026	0.132	7/10	24.48
70	1 500	0.000	0.000	0.000	10/10	12.62	0.031	0.123	0.318	0/10	30.40
80	1 700	0.000	0.011	0.054	8/10	30.01	0.000	0.088	0.116	2/10	38.80
90	1 850	0.000	0.158	0.446	2/6	13.83	0.000	0.012	0.061	8/10	39.26
100	2 000	0.000	0.000	0.000	6/6	2.50	0.000	0.115	0.417	4/10	35.01
总计		0.000	0.019	0.446	60/78	7.23	0.003	0.047	0.417	51/90	26.00

注: min 为所得解的目标值  $Z$  与  $Z_{better}$  的误差的最小值, ave 为所得解的目标值  $Z$  与  $Z_{better}$  的误差的平均值, max 为所得解的目标值  $Z$  与  $Z_{better}$  的误差的最大值,  $\frac{BK}{\text{Feasible}}$  为求得目标值为  $Z_{better}$  的次数与得到可行解的次数的比, Time 为平均运行时间。

假定需求点数目  $|I| = 20, T = 0.5$ , 当预算  $B$  变化时, 同时使用两类启发式算法得到  $Z_{better}$  的结果如图 2 所示。图 2(a) 表示随着预算的增加, 服务的质量提高, 总体上目标值越大, 也就是在承诺的最大时间内完成服务的平均需求率越多。图 2(b) 显示随着预算增加, 平均每个需求的平均运输时间总体趋势为递减。平均运输时间增加表示服务台相对集中, 选择的设施点较少, 此时订单在系统中的平均逗留时间减少; 平均运输时间减少表示服务台相对分散, 选择的设施点较多, 每个订单在系统中的平均逗留时间增加。图 2(b) 显示平均运输时间与平均逗留时间的波动基本上相反。随着预算增加, 平均每个需求的平均响应时间基本上减少。

假定  $B = 600, T = 0.5$ , 当需求点  $|I|$  变化时, 同时使用两类启发式算法得到  $Z_{better}$  的结果如图 3 所示。随着需求点的增加, 目标值  $Z_{better}$  总体趋势为先增加后减少。随着需求增加, 在承诺时间内完成的服务的

百分比  $P$  减少, 开始时随着需求增加,  $P$  减少不明显, 目标值总体趋势为增加, 当  $|I|$  达到 30 左右时, 目标值达到最大, 随着  $|I|$  增大,  $P$  剧减, 使目标值减少。随着需求点的增加, 平均每个需求的平均运输时间总体趋势为增加, 造成平均每个需求的平均响应时间总体趋势也为增加。

假定  $B = 600, |I| = 30$ , 随着  $T$  的变化, 同时使用两类启发式算法得到  $Z_{better}$  的结果如图 4 所示。随着  $T$  的增加,  $Z_{better}$  明显增加, 因为在最大响应时间内完成服务的百分比  $P$  增加, 自然在最大响应时间  $T$  内完成服务的平均需求增加。图 4(b) 中显示的平均每个需求的平均逗留时间和平均运输时间的波动基本上是相反。

### 6 结论

本研究探讨带有响应时间承诺的设施的选址问题, 证明系统逗留时间分布函数是关于服务台数目

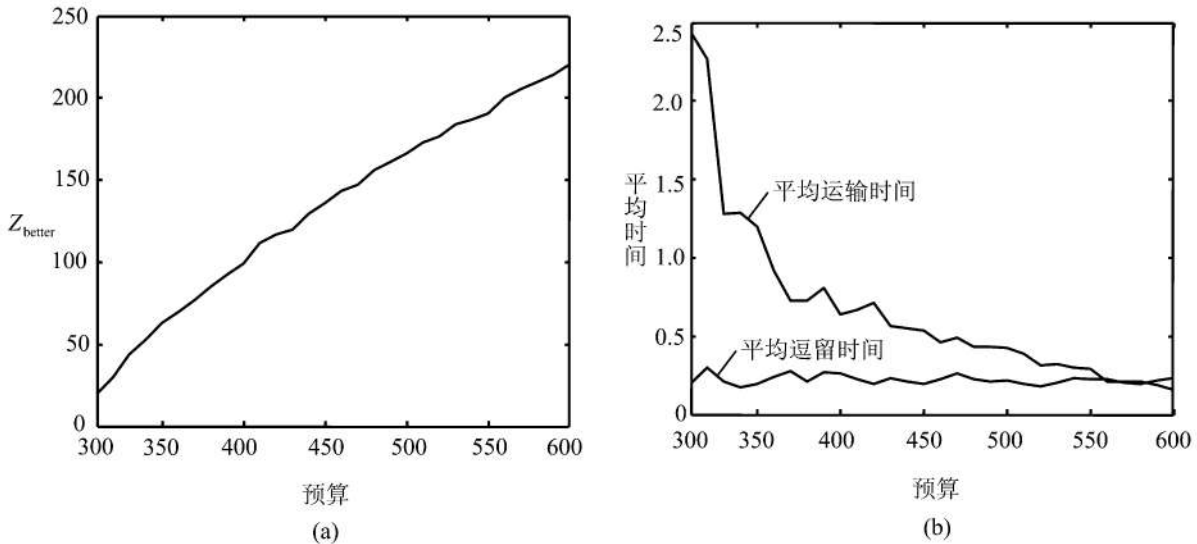


图 2 不同预算的结果分析  
Figure 2 The Results of Different Budgets

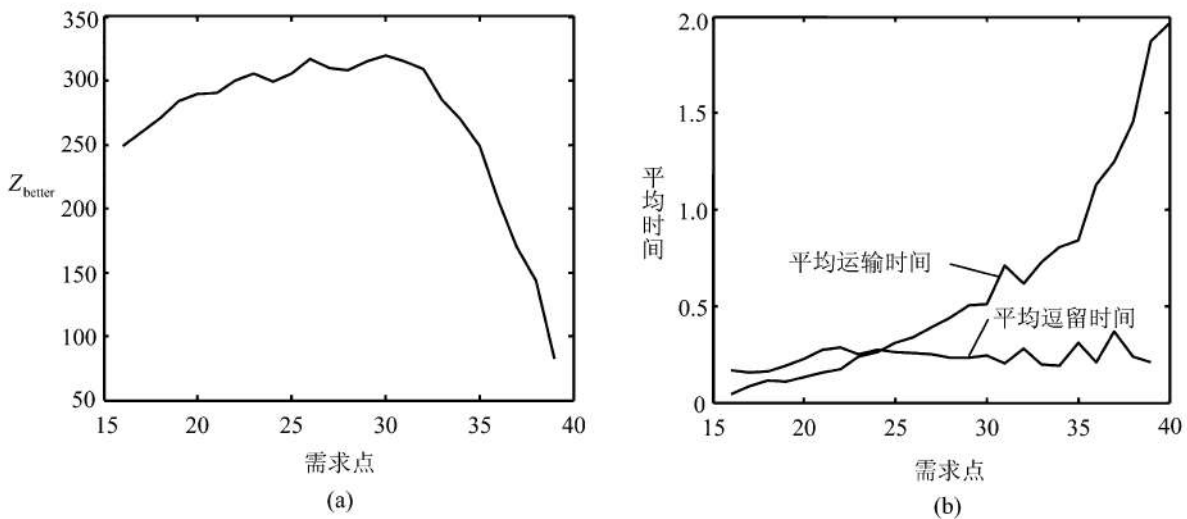


图 3 不同需求点数量的结果分析  
Figure 3 The Results of Different Quantity of Demand Nodes

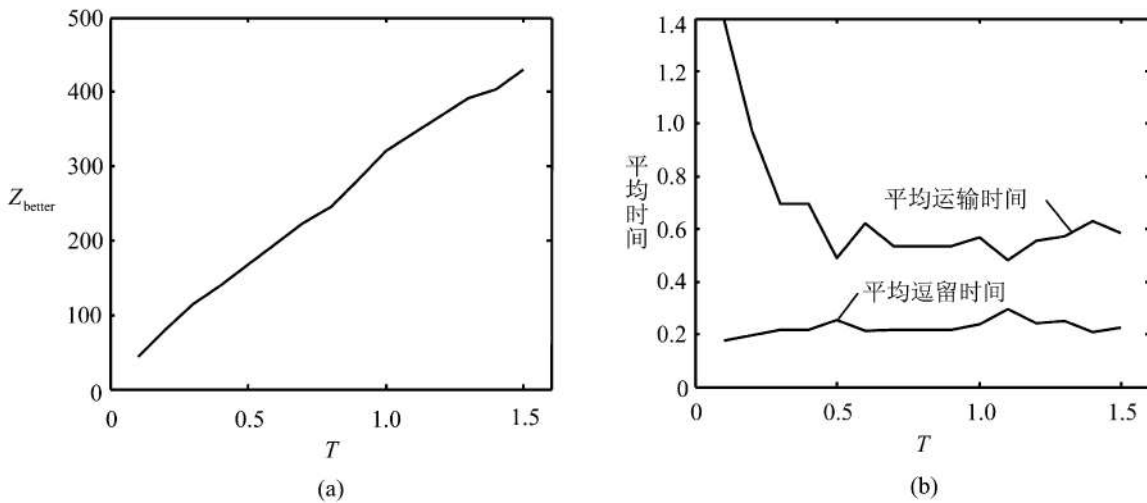


图4 不同  $T$  值的结果分析  
Figure 4 The Results of Different  $T$

的凹函数,并由此给出最优服务台分配算法,同时给出两类启发式算法,即下降算法和遗传算法,通过随机算例验证,计算结果较满意,并对不同的参数进行分析。研究表明,在承诺的响应时间内得到满足的平均需求数量随着预算和承诺时间的增加而增加,而平均需求数量是呈现先增加后减少的趋势。这个结果与现实情况是相符的,预算的增加会使服务台或设施数量增加,从而缩短订单在系统中的平均排队等待时间或是平均运输时间,使平均每个需求的平均响应时间减少,承诺时间的增加无疑也会使满足该响应时间的需求数量增加。需求数量变化时目标函数变化情况并不是呈单调趋势,当需求数量较少、服务台不能充分利用时,在承诺的响应时间内得到服务的平均需求数量随着总需求数量的增加而增加;当需求数量过多、服务台工作繁忙时,总需求数量的增加反而会使达到满意服务的平均需求数量减少。数值模拟结果还表明,随着参数的变化平均每个需求的平均逗留时间和平均运输时间的波动基本上是相反的,通过平衡这两段时间,使在承诺的最大响应时间内完成服务的平均需求数量最多。

未来研究可考虑服务时间和需求产生为其他随机分布的情况;本研究中的设施相互之间互相独立,还可以研究设施之间相互合作的拥塞选址;在遇到设施拥塞时,考虑部分顾客具有优先权的选址问题;考虑多种类顾客、多元化服务排队选址问题;在供应链中的排队选址研究需要考虑订货策略、库存情况、多级供应链等更为复杂的情况。

#### 参考文献:

- [1] Daskin M S. Network and Discrete Location: Models, Algorithms and Applications [M]. New York: John Wiley and Sons, 1995.
- [2] Boffey B, Galvao R, Espejo L. A Review of Congestion Models in the Location of Facilities with Immobile Servers [J]. European Journal of Operational Research, 2007, 178(3): 643-662.
- [3] Berman O, Larson R C, Chiu S. Optimal Server Location on a Network Operating as an M/G/1 Queue [J]. Operations Research, 1985, 33(4): 746-771.
- [4] Brimberg J, Mehrez A, Wesolowsky G O. Allocation of Queuing Facilities Using a Minimax Criterion [J]. Location Science, 1997, 5(2): 89-101.
- [5] Scott C, Jefferson T, Drenzer Z. Various Objectives for the Queueing-location Problem on the Plane [J]. Asia-Pacific Journal of Operational Research, 1999, 16(2): 203-214.
- [6] Silva F, Serra D. Incorporating Waiting Time in Competitive Location Models [J]. Networks & Spatial Economics, 2007, 7(1): 63-76.
- [7] Wang Q, Batta R, Rump C. Algorithms for a Facility Location Problem with Stochastic Customer Demand and Immobile Servers [J]. Annals of Operations Research, 2002, 111(1): 17-34.
- [8] Zhang L, Rushton G. Optimizing the Size and Locations of Facilities in Competitive Multi-site Service Systems [J]. Computers & Operations Research, 2008, 35(2): 327-338.
- [9] Marianov V, Ríos M, Icaza M J. Facility Location for Market Capture When Users Rank Facilities by Shorter Travel and Waiting Times [J]. European Journal of Operational Research, 2008, 191(1): 32-44.
- [10] Benjaafar S, Li Y, Xu D, Elhedhli S. Demand Allocation in Systems with Multiple Inventory Locations and Multiple Demand Sources [J]. Manufacturing & Service Operations Management, 2008, 10(1): 43-60.
- [11] Huang S, Batta R, Nagi R. Distribution Network Design: Selection and Sizing of Congested Connections

- [ J ]. *Naval Research Logistics*, 2005, 52(8): 701–712.
- [12] Lieckens K, Vandaele N. Reverse Logistics Network Design with Stochastic Lead Times [ J ]. *Computers & Operations Research*, 2007, 34(2): 395–416.
- [13] Morabito R, Chiyoshi F, Galvao R D. Non-homogeneous Servers in Emergency Medical Systems; Practical Applications Using the Hypercube Queueing Model [ J ]. *Socio-Economic Planning Sciences*, 2008, 42(4): 255–270.
- [14] Iannoni A P, Morabito R, Saydam C. A Hypercube Queueing Model Embedded into a Genetic Algorithm for Ambulance Deployment on Highways [ J ]. *Annals of Operations Research*, 2008, 157(1): 207–224.
- [15] Marianov V, Serra D. Probabilistic, Maximal Covering Location-allocation Models for Congested System [ J ]. *Journal of Regional Science*, 1998, 38(3): 401–424.
- [16] Shavandi H, Mahlooji H. A Fuzzy Queueing Location Model with a Genetic Algorithm for Congested Systems [ J ]. *Applied Mathematics and Computation*, 2006, 181(1): 440–456.
- [17] Berman O, Krass D, Wang J. Locating Service Facilities to Reduce Lost Demand [ J ]. *IEEE Transaction*, 2006, 38(11): 933–946.
- [18] Berman O, Drezner Z. The Multiple Server Location Problem [ J ]. *Journal of the Operational Research Society*, 2007, 58(1): 91–99.
- [19] Aboolian R, Berman O, Drezner Z. The Multiple Server Center Location Problem [ J ]. *Annals of Operations Research*, 2009, 167(1): 337–352.
- [20] 郭占宽, 常利胜, 尹健, 刘建芳. 一类排队系统等待时间的分布 [ J ]. *内蒙古大学学报(自然科学版)*, 2008, 39(4): 375–379.  
Guo Z K, Chang L S, Yin J, Liu J F. The Distribution of Waiting Time in a Class of Queueing Systems [ J ]. *Journal of Inner Mongolia University (Acta Scientiarum Naturalium Universitatis Neimongol)*, 2008, 39(4): 375–379. (in Chinese)
- [21] Jagers A A, VanDoorn E A. Convexity of Function Which Are Generalizations of the Erlang Loss Function and the Erlang Delay Function [ J ]. *SIAM Review*, 1991, 33(2): 281–282.

## Servers Location-allocation Problem with Promise of Response Time

HU Dan-dan<sup>1</sup>, YANG Chao<sup>1</sup>, LIU Zhi-wei<sup>2</sup>

1 School of Management, Huazhong University of Science and Technology, Wuhan 430074, China

2 Control Science & Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

**Abstract:** In fierce competition environment, improving service quality becomes an effective way of winning for an enterprise, among which, response time is one of vital factors to evaluate service quality. Many facilities make promise of the maximum response time for customers, especially in service business. Much attention has been focused on the need to location facilities for attracting more customers and obtaining higher reputation. To make decisions of these kinds of facilities locations not only simply consider the region that can the facilities service cover, but also the service time, which should be less than the promised maximum response time. A nonlinear integer programming model with budget constrained is built through combining location theory with queue theory organically. The objective of the model is to figure out proper locations and relative server number in order to maximize the demand quantity within the promised maximum response time which includes the order's sojourn time at facilities and the transport time. The results prove that sojourn time distribution function in system is a concave one of the server number. Meanwhile, heuristic algorithms are provided. Through plentiful random samples, the conclusion that the algorithms are effective is given.

**Keywords:** location; queueing; multi-server; heuristic algorithm

**Received Date:** February 24<sup>th</sup>, 2009      **Accepted Date:** October 26<sup>th</sup>, 2009

**Funded Project:** Supported by the National Natural Science Foundation of China(70601011, 70871044) and the Trans-Century Training Program Foundation for the Talents by the State Education Commission (NCET-06-0653).

**Biography:** HU Dan-dan, a Zhejiang Cixi native(1983 -), is a Ph. D. candidate in the School of Management at Huazhong University of Science and Technology. Her research interests include network optimization and decision optimization, etc. E-mail: hudandan1224@163.com      □