

机场出租车载客优化问题研究

张楚悦, 黄明源, 闫栋栋, 伍嘉琪, 徐晓凤, 孙菊贺

沈阳航空航天大学理学院, 辽宁 沈阳

收稿日期: 2022年10月28日; 录用日期: 2022年11月23日; 发布日期: 2022年11月30日

摘要

针对机场出租车司机等待载客、放空返回的选择问题, 研究了一种提高出租车司机运载乘客效率与收益的选择策略。根据模型的合理性和相关因素的依赖性, 建立了基于多属性的出租车司机接客决策树模型, 并通过Python的爬虫技术获得交通信息数据集, 在基于大数据的程序批量预处理后, 协同其他不确定属性验证模型的可行性。建立了数学模型, 试验测试计算结果表明: 当计算得平均到达率 $\lambda = 5.98$ 、单个上车点的平均服务率 $\mu = 2.21$ 、乘客等待的时间费用 $w = 0.001$ 时, 模型可靠性最高, 且上车点数量 m 等于5时乘客时间成本、出租车时间成本与效益值达到最优, 实现排队服务系统成本和费用最小。

关键词

决策树模型, 费用决策, 大数据处理, Python爬虫

Research on Optimization of Airport Taxi Carrying Passengers

Chuyue Zhang, Mingyuan Huang, Dongdong Yan, Jiaqi Wu, Xiaofeng Xu, Juhe Sun

School of Science, Shenyang Aerospace University, Shenyang Liaoning

Received: Oct. 28th, 2022; accepted: Nov. 23rd, 2022; published: Nov. 30th, 2022

Abstract

Aiming at the choice problem of airport taxi drivers waiting for passengers and returning empty, a selection strategy to improve the efficiency and income of taxi drivers to carry passengers is studied. According to the rationality of the model and the dependence of related factors, a multi-attribute-based taxi driver pick-up decision tree model is established, and a traffic information dataset is obtained through Python crawler technology. After batch preprocessing of programs based on big data, the feasibility of other uncertain attribute verification models is coordinated. A mathematical model is established, and the experimental test calculation results show that when the cal-

culated average arrival rate $\lambda = 5.98$, the average service rate $\mu = 2.21$ of a single boarding point, and the passenger waiting time cost $w = 0.001$, the reliability of the model is the highest, and when the number of boarding points m is equal to 5, the passenger time cost, taxi time cost and benefit values are optimal, and the cost and cost of the queuing service system are minimized.

Keywords

Decision Tree Model, Cost Decision, Big Data Processing, Python Crawler

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,随着世界经济的飞速发展,人民生活水平不断提高,旅游业也得到了前所未有的发展,所以无论是从频率还是从出行方式上,人们的选择意识都发生了很大转变,飞机已经逐渐成为人们长途旅行的主流选择。由于机场大多是修建在远离市区的地方,位置相对比较偏,很多乘客下飞机后要去市区或周边的目的地,出租车是机场出行的主要交通工具。

出租车作为乘客出行方式的重要一环[1],在其中充当着重要作用,并且多年来形成的传统的运输运营模式已经无法满足旅客、司机等各方面的需求[2]。如何进行调配以平衡出租车与乘客之间的供需关系[3];以及怎么抉择,保证司机利益最大化的同时[4],尽可能满足乘客需求,在供应乘客运输的条件下[5],最大程度提高出租车运输效率是摆在机场管理方面的一道难题[6],受到许多学者们的关注[7]。更进一步的,搭载乘客的人数和行驶里程与出租车司机的收益都是有相关性的[8],短途载客的司机与长途载客的司机、满载乘客的司机与载乘客较少的司机相比较而言,利益差距过大。为了平衡出租车司机的收益,管理部门欲给一部分短途载客并且继续返回机场拉客的出租车一定的“优先权”,而给长途载客并且需要稍等待乘客满载的司机一定的“滞留”时间,从而需要给出一个可以在现实中实行的优先安排方案。针对乘客与出租车供求关系不平衡所导致的乘车效率低下的问题,我们将问题模型化,通过建立合适的上车点、乘车时间段,对出租车与乘客合理调度,使得总的乘车效率提高。首先是对已有的相关研究进行回顾,其次针对不同的问题建立最优化模型,之后讨论所建模型中待求参数的估计方法与结果,最后对模型的求解结果进行检验并总结。

2. 机场出租车问题研究

2.1. 决策树模型

对于出租车司机选择空载返回还是选择等待载客的决策问题,本文选择用决策树模型进行求解。

定义熵的概率分布为 $P(X = x_i) = p_i, i = 1, 2, \dots, n$, 那么将出租车司机决策模型中特征 A 给定条件 D 的经验条件熵 $H(D)$ 与特征 A 给定条件下 D 的经验条件熵 $H(D|A)$ 之差记为 $g(D|A)$, 其中 $g(D, A) = H(D) - H(D|A)$ 。根据决策树模型理论,可以得到如下结论:

1) 数据集 D 的经验熵:

$$H(D) = -\sum_{k=1}^K \frac{|C_k|}{|D|} \log_2 \frac{|C_k|}{|D|} \quad (1)$$

2) 特征值 A 对数据集 D 的经验条件熵:

$$H(D|A) = \sum_{i=1}^n \frac{|D_i|}{|D|} H(D_i) = -\sum_{i=0}^n \frac{|D_i|}{|D|} \sum_{k=1}^K \frac{|D_{ik}|}{|D_i|} \log_2 \frac{|D_{ik}|}{|D_i|} \quad (2)$$

3) 信息增益:

$$g(D, A) = H(D) - H(D, A) \quad (3)$$

4) 特征 A 对训练数据集 D 的信息增益比:

$$g_k(D, A) = \frac{g(D, A)}{H(D)} \quad (4)$$

根据对上述决策树的信息增益分析, 绘制出的决策树模型如图 1 所示, 其中 Z 表示判断结果, Qty 表示排队等候收益, Qtn 表示空载返回收益, t_1 表示排队等候到接到乘客的时间, t_3 表示空载返回到接到乘客的时间, Tp 表示乘客的疏散时间, Tc 表示出租车的疏散时间。

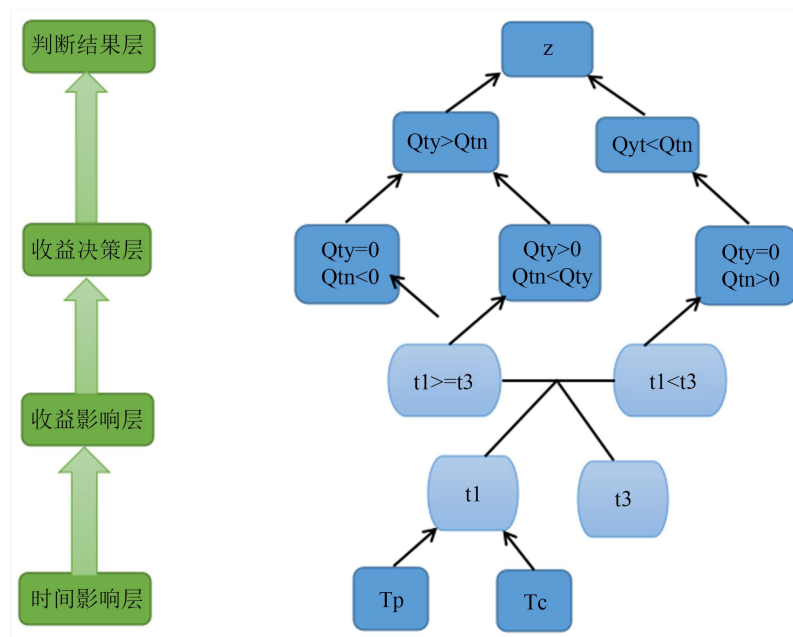


Figure 1. Flowchart of decision tree model

图 1. 决策树模型流程图

2.2. Python 爬虫

使用实际机场数据对 2.1 中的决策树模型加以验证, 数据收集及预处理的过程如下:

1) 先基于决策树模型的底层影响因素, 判断出影响司机决策的主观因素需要用到的验证集, 即该时间段航班数据集和最高效空载距离集;

2) 然后通过 Python 爬虫技术, 获得沈阳桃仙机场入港航班的数据集和沈阳出租车 GPS 定位的打车需求量和出租车分布的特征数据集;

3) 再通过 Python 进行数据集清洗预处理, 得到某一时间段航班数量表和最高效空载距离, 通过类比的推理的思想, 将百度地图热度图数据作为短距离运输最高效空载距离参考;

4) 最后通过 Python 进行决策树构建以及大数据运算, 得出每天每个时间段的出租车司机决策结果。

2.3. 费用决策

若要在供应乘客运输的条件下, 最大程度提高出租车运输效率, 这要求解使乘客的等待时间和上车点建设成本最少的上车点个数, 使得总的乘车效率最高。很显然, 上面所描述的是一个带有约束条件的最优化问题。

本文设计一个预方案以预估结果: 假设第一个出租车在第一个上车点接人, 从第一个出租车停在第一个上车点的那一时刻开始计时, 即此时 $t = 0$, 后面来的每一辆车都找到最近的且没有其他出租车停靠的上车点停车接客, 将上车点依次编号为①, ②, ③, ④, ⑤……, 再把按顺序进入乘车区的批量出租车的编号为 $i = 1, \dots, n$ 。下面列举了 $n = 2, 4, 8$ 时可以设置的上车点的最少数量以及乘客分布, 数据如表 2、表 3、表 4 所示。其中 C_i 表示第 i 量出租车在某时刻, 停留在相应编号的上车点。

Table 1. $n = 2$ the minimum number of pick-up points and the distribution of passengers

表 1. $n = 2$ 时上车点的最少数量以及乘客分布

	①	②	③	④
$t = 0$	C1	0	0	0
$t = 0.5$	C1	0	0	0
$t = 1$	C1	0	0	0
$t = 1.5$	C4	C2	0	0
$t = 2$	C4	C2	C3	0
$t = 2.5$	C4	C2	C3	0
$t = 3$	0	0	C3	0

Table 2. $n = 4$ the minimum number of pick-up points and the distribution of passengers

表 2. $n = 4$ 时上车点的最少数量以及乘客分布

	①	②	③	④
$t = 0$	C1	0	0	0
$t = 0.5$	C1	0	0	0
$t = 1$	C1	0	0	0
$t = 1.5$	C4	C2	0	0
$t = 2$	C4	C2	C3	0
$t = 2.5$	C4	C2	C3	0
$t = 3$	0	0	C3	0

Table 3. $n = 8$ the minimum number of pick-up points and the distribution of passengers

表 3. $N = 8$ 时上车点的最少数量以及乘客分布

	①	②	③	④
$t = 0$	C1	0	0	0
$t = 0.5$	C1	0	0	0
$t = 1$	C1	0	0	0
$t = 1.5$	C4	C2	0	0
$t = 2$	C4	C2	C3	0
$t = 2.5$	C4	C2	C3	0
$t = 3$	C7	C5	C3	0
$t = 3.5$	C7	C5	0	0

Continued

$t = 4$	C7	C5	0	0
$t = 4.5$	0	C8	C6	0
$t = 5$	0	C8	C6	0
$t = 5.5$	0	C8	C6	0

从表 1~3 可以得到的结论是：当 $n = 2$ 时，至少设置两个上车点，每个上车点乘客比例约为 1:1；当 $n = 4$ 时，至少设置三个上车点，每个上车点乘客比例约为 2:1:1；当 $n = 8$ 时，至少设置三个上车点，每个上车点乘客比例约为 3:3:2。

下面令 λ 表示乘客的平均到达率， μ 表示单个上点的平均服务率， m 表示上车点的个数， η 表示系统服务率， w 表示乘客等待的时间费用， L_y 表示排队服务系统决策成本费用。由于机场外的行车道不会无限长，所以排队系统在稳定工作状态时，系统中出租车乘客数概率如下：

$$P_0(m) = \left[\sum_{i=0}^{m-1} \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k + \frac{1}{m!} \frac{1}{(1-\eta)} \left(\frac{\lambda}{\mu}\right)^m \right]^{-1} \tag{5}$$

$$P_n(m) = \begin{cases} \frac{1}{n} \left(\frac{\lambda}{\mu}\right)^k P_0(m), n = 1, 2, \dots, m \\ \frac{1}{m! m^{n-m}} \left(\frac{\lambda}{\mu}\right)^n P_0(m), n = m + 1 \end{cases} \tag{6}$$

对系统进行分析得：

$$L_y = L_x + m\eta = \frac{1}{m!} \frac{(m\eta)^m \eta}{(1-\eta)^2} P_0 + \frac{\lambda}{\mu} \tag{7}$$

$$E(T) = \frac{P_n(m)}{m\mu(1-\eta)^2} = \frac{n\mu}{n!(n\mu - \lambda)^2} \left(\frac{\lambda}{\mu}\right)^n P_0(m) \tag{8}$$

利用排队系统的费用决策模型对排队服务系统进行优化，需要满足两者之和最小才能使排队系统进一步优化，即：

$$\min Z(m) = Z_1 + Z_2 \tag{9}$$

$$Z(m) \leq Z(m+1) \tag{10}$$

$$Z(m) \leq Z(m-1) \tag{11}$$

其中，约束条件为：

$$\omega(L_y(m) - L_y(m-1)) \leq \nu \leq \omega(L_y(m+1) - L_y(m)) \tag{12}$$

根据搭乘出租车的乘客数可计算出对应的乘客的到达率和系统服务率，结果如表 4 所示。

Table 4. Passenger arrival rate and service rate

表 4. 乘客到达率与服务率

λ/min	μ/min	ω/ν
6	2	0.001

接下来使用费用决策模型对沈阳桃仙国际机场的出租车上车点及排队服务系统进行优化，费用决策模型计算结果如表 5 所示。

Table 5. Calculation results of the cost decision model
表 5. 费用决策模型计算结果

$L_y(1) - L_y(2)$	11.239	$L_y(5) - L_y(6)$	2333.36128
$L_y(2) - L_y(3)$	-16.994	$L_y(6) - L_y(7)$	3279.55554
$L_y(3) - L_y(4)$	-29.94		
$L_y(4) - L_y(5)$	-668.32		

由上述计算结果可知，当 $m = 5$ 时，满足原模型的约束条件，也即此时乘客时间消费和上车点及系统的成本之和最小。这说明了当设置 5 个出租车上车点时，可实现排队服务系统的成本和费用相对较小。

2.4. 应用

2.4.1. 应用说明

为了平衡出租车司机的收益，需要给出一个“优先安排”方案。这可以概括为需要对之前建立的决策树模型的时间距离收益属性加以改善，提供有针对性的补偿措施。

首先将基于 GPS 定位的打车需求量和出租车分布的特征数据集进行目的地概率预测，得出地点概率预测列表和距离概率期望值；其次求出决策树模型第二层的时间距离收益的拐点，找到即使比排队时早接客但是由于空载原因使收益仍然低于排队的时间或距离区间；再对这个区间进行长途短途分类；最后结合实际情况得出最优方案。

2.4.2. 实例验证

如表 6 所示，我们将沈阳桃仙机场 2021 年 9 月 12 日出租车进场出场数据作为初始条件，以 1 小时为时间间隔，计算 24 个时间段内各自所需的最少上车点。

首先，利用 2.3 费用决策中的预方案，分别计算 $n = 2$ 、 $n = 4$ 、 $n = 8$ 时上车点的最少数量，如表 7 所示；根据费用决策模型的约束条件，计算得出乘客平均到达率 λ 、单个上车点的平均服务率 μ 、乘客等待的时间费用 w 分别为 5.98、2.12、0.001，与费用决策模型计算结果的相对误差较小；经上述数据及相关计算得到了如表 8 所示的排队服务系统决策成本费用表，由此可知，当上车点数量 m 等于 5 时，乘客时间成本、出租车时间成本与效益值达到最优，实现排队服务系统成本和费用较小。

Table 6. Taoxian airport 2021.9.12 initial data
表 6. 桃仙机场 2021.9.12 初始数据

时间段	场内待运车辆数	进场车辆数	出场车辆数
0:00~1:00	19	42	45
1:00~2:00	15	38	40
2:00~3:00	25	52	48
3:00~4:00	24	53	55
4:00~5:00	18	42	40
5:00~6:00	19	39	40
6:00~7:00	19	41	40
7:00~8:00	20	43	40

Continued

8:00~9:00	38	74	77
9:00~10:00	52	99	95
10:00~11:00	45	103	95
11:00~12:00	31	69	71
12:00~13:00	25	52	49
13:00~14:00	29	55	50
14:00~15:00	28	60	61
15:00~16:00	33	70	71
16:00~17:00	39	75	81
17:00~18:00	46	92	95
18:00~19:00	51	101	93
19:00~20:00	47	92	93
20:00~21:00	51	98	95
21:00~22:00	34	68	70
22:00~23:00	24	50	54
23:00~24:00	28	57	54

Table 7. $n = 2$, $n = 4$, and $n = 8$ the minimum number of pick-up points in the expense decision表 7. 费用决策中 $n = 2$ 、 $n = 4$ 、 $n = 8$ 时上车点的最少数量

时间段	$n = 2$	$n = 4$	$n = 8$
0:00~1:00	2	5	5
1:00~2:00	2	4	5
2:00~3:00	3	5	5
3:00~4:00	3	5	5
4:00~5:00	2	5	5
5:00~6:00	2	4	5
6:00~7:00	2	5	5
7:00~8:00	2	5	5
8:00~9:00	4	5	5
9:00~10:00	5	5	5
10:00~11:00	5	5	5
11:00~12:00	4	5	5
12:00~13:00	3	5	5
13:00~14:00	3	5	5
14:00~15:00	3	5	5
15:00~16:00	4	5	5
16:00~17:00	4	5	5
17:00~18:00	5	5	5
18:00~19:00	5	5	5
19:00~20:00	5	5	5

Continued

20:00~21:00	5	5	5
21:00~22:00	4	5	5
22:00~23:00	3	5	5
23:00~24:00	3	5	5

Table 8. Schedule of decision costs for queuing service systems

表 8. 排队服务系统决策成本费用表

$L_y(1) - L_y(2)$	12.396
$L_y(2) - L_y(3)$	-10.284
$L_y(3) - L_y(4)$	324.179
$L_y(4) - L_y(5)$	3894.652

3. 模型评估

3.1. 模型的优缺点

1) 模型的优点

a) 决策树模型能够将一系列大数据进行处理,从而得到可靠结果;费用决策模型将抽象问题具象化,使模型更易于理解;

b) 设置多组变量分别对可能产生的多种因素具体讨论,增强模型合理性;

c) 模型比较简单,能够直观地反映出权重的相关影响因素以及权重之间的关系;

d) 模型假设不考虑微观因素的影响,使得模型更易操控,普遍性和适用性更强。

2) 模型的缺点

a) 本文所建立的模型对数据的依赖性很强,必须保证数据本身的准确性;

b) 多点纵列式排队模型对区域的纵向距离要求较高,一般适用于机场交通枢纽,道路较短的不适用;

c) 模型设计中考虑的情况不够完全,假设过于理想化;

d) 模型拓展性不高,容易受外界因素的影响。

3.2. 模型的推广与改进

本文所建立的优化模型除了适用于机场的出租车问题外,还适用于客运站拉客等一系列接客问题。虽然本文提出出租车停靠方案能提高乘客的乘车效率,但是影响乘车效率的因素还有很多,比如天气、时间等,因此还可以将更多因素考虑进去对模型加以改进。

参考文献

- [1] 魏中华,王琳,邱实.基于排队论的枢纽内出租车上客区服务台优化[J].公路交通科技(应用技术版),2017,13(10):298-300.
- [2] 黄岩,王光裕.虹桥机场T2航站楼出租车上客系统组织管理优化探讨[J].城市道桥与防洪,2014(12):7-9+35-36.
- [3] 林思睿.机场出租车运力需求预测技术研究[D]:[硕士学位论文].成都:电子科技大学,2018.
- [4] 商瑞杰,周大勇,程扬,张硕.多属性决策模型在机场出租车司机决策问题上的应用[J].黑龙江科学,2021,12(22):8-11.
- [5] 魏逸凡.机场出租车司机决策问题研究[J].数码世界,2020(4):193.

- [6] 徐小博, 谢振炎, 程雅楠. 基于运筹学原理研究机场出租车排队优化问题[J]. 内蒙古教育, 2020(33): 60-72.
- [7] 廖周宇, 程艺明, 石敏斐. 机场出租车智能调度策略研究[J]. 科技视界, 2020(19): 240-243.
- [8] 朱桂玲. 机场出租车最优化模型调配问题的研究[J]. 商讯, 2019(32): 179-180.