网络出版时间：2019-11-20 14:15:57

第 32 卷　第 4 期　　石家庄铁道大学学报(自然科学版)　　Vol. 32　No. 4
2019 年 12 月　Journal of Shijiazhuang Tiedao University(Natural Science Edition)　　Dec. 2019

出租车空车率影响因素研究

唐隽玉，　朱　祎，　黄一哲

（上海交通大学 船舶海洋与建筑工程学院，上海　200240）

摘要：出租车是都市交通中必不可少的一部分。其空车率水平越发受到司机、乘客以及管理者的关注。全面理解空车率的影响因素对平衡出租车运行效率和乘客等待时间之间的矛盾至关重要。基于出租车运行全轨迹，旨在挖掘空车率的影响因素并量化空车率和影响因素之间的关系。为区分不同的空车率，先将空车率分为高、中、低 3 种水平。再分别分析空/重车行程，从寻/送客策略的角度提取出了 5 个可能的影响因素。最后，构建广义多水平定序 Logit 模型以识别显著影响因素。结果表明，这 5 个因素对出租车空车率影响显著，并且各因素对不同空车率水平的出现概率影响不同，为调整空车率水平和优化出租车运营管理提供了指导。

关键词：出租车；空车率；GPS；定序 Logit 模型；影响因素

中图分类号：U121　**文献标志码**：A　**文章编号**：2095 - 0373(2019)04 - 0097 - 06

0　引言

出租车是日常生活中不可或缺的交通工具，为人们提供便捷的门到门运输服务。出租车供需的随机性，导致了出租车空车率水平的不合理问题。空车率过高会增加司机的工作成本，造成时间和燃料的浪费。而空车率过低会延长乘客的等待时间。此外，过多的空车将会加剧交通拥堵，并造成空气污染，如在台湾，空车每年会造成 9 000 万 L 汽油的浪费[1]。为缓解上述问题，一些学者对出租车空车时长/空车率进行了一系列相关研究。关金平等[2]分析了出租车空驶的时空特性，并且从人文地理、城市规划角度分析了成因。鞠炜奇等[3]以深圳为例对出租车空车率的时空分布特征及影响因素进行了分析。但上述两项关于出租车影响因素的研究仅停留在定性分析阶段。量化空车率和影响因素之间的关系，可为调节空车率至合理水平提供科学依据。提取空车率影响因素是量化二者关系的第一步，现有研究表明以下因素与出租车的空车率有密切关联：寻/送客时长、距离[4]，司机寻客策略[5]，上车次数[6]等。然而，目前尚缺少将这些因素综合起来进行研究的文献。

Logit 模型作为一种数学工具被广泛应用于城市交通研究中[7]，本文基于广义多水平定序 Logit 模型(GMOL 模型)，旨在从驾驶员行为分析角度，建立一个全面且定量化的方法来挖掘出租车空车率的影响因素，从而寻求影响因素与空车率之间的量化关系，以合理化调节空车率水平，达到优化出租车运行效率及乘客满意度的目的。

1　数据描述及预处理

1.1　数据描述与清洗

GPS 数据由上海强生出租车公司采集，涵盖 10 000 辆以上的出租车运行信息，平均每 10 s 记录一次。每条记录包括出租车 ID 号(唯一标记)、载客状态(1 表示空车，0 表示重车)、GPS 信息接收时间、当前位置的经、纬度以及瞬时速度。

收稿日期：2018 - 01 - 24　**网络出版日期**：--　**责任编辑**：车轩玉　DOI：10. 13319/j. cnki. sjztddxxbzrb. 20180028

网络出版地址：http://kns.cnki.net/kcms/detail/13.1402.n.20191120.1354.016.html

基金项目：国家社会科学基金(41701552)

作者简介：唐隽玉(1993—)，女，硕士研究生，主要从事交通大数据分析研究。E-mail：tjysghr@sjtu. edu. cn

唐隽玉，朱祎，黄一哲. 出租车空车率影响因素研究[J]. 石家庄铁道大学学报：自然科学版，2019，32(4)：97-102.

由于GPS信号遮挡、设备故障等原因,需要进行数据清洗。将经纬度在[120.852°E, 121.925°E],[30.693°N, 31.511°N]之外、瞬时速度在0,120 km/h之外的数据进行剔除,删除了占原始数据0.007%的异常数据。

1.2 时空划分

对于空间划分,本文的研究区域为除去崇明岛的上海市主干区域,并将研究区域网格化,即将上海主干区域划分为一系列约为500 m×400 m大小相同的网格,总量为22 814个。

对于时间划分,由于周五相较于其它4个工作日呈现出不同的出租车驾驶模式,同时为了减少计算复杂度,选择2016年3月21日—2016年3月24日(周一至周四)作为计算原始数据。此外,还需对研究时段进行划分。

如图1所示,载客车速度作为筛选研究时段的第一个指标,如果速度过低,说明当前路况拥挤,司机不能自主地采取策略进行运营。而且,为保证足够的样本量,运营车数量作为第二个筛选时间段的指标。最后,还需要排除司机用餐时间的影响,Qin et al[8]对上海市出租车司机的用餐时间进行调研,发现用餐时间灵活地分布在11:00~14:00以及16:00~19:30之间。综上,选取14:00~16:00作为研究时间段。

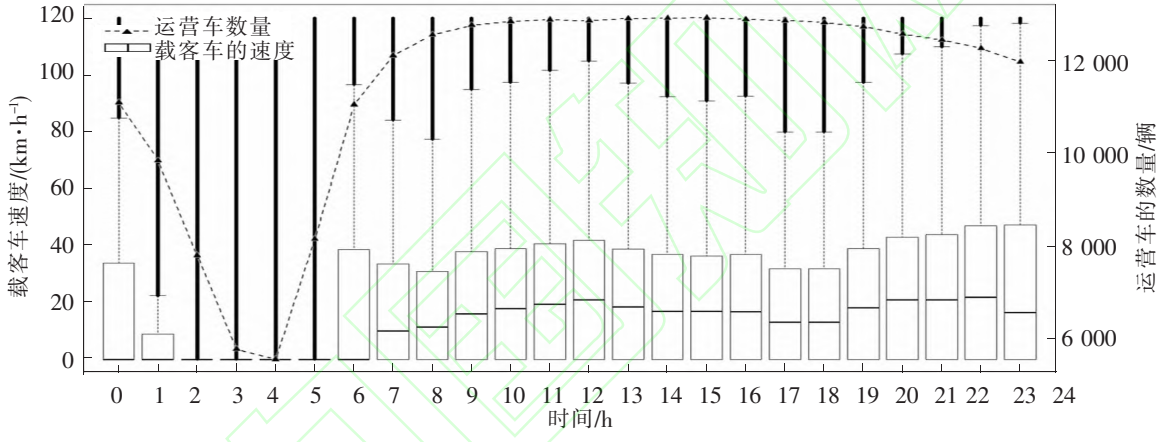

图1 载客车速度、运营车数量在一天中的变化

2 空车率与影响因素量化模型

2.1 出租车空车率的定义

视出租车的时间空车率为空车率的衡量标准,因为出租车的机会成本是通过空车时间测算而非空车运营距离测算[9]。司机i的空车率VR_i的计算如下

$$VR_i=\frac{\sum_j t_{i,j}^1}{\sum_j t_{i,j}^1+\sum_j t_{i,j}^0} \tag{1}$$

式中,$t_{i,j}^0$对应司机i的第j次重车行程的运营时间;$t_{i,j}^1$对应司机i的第j次空车行程的运营时间。

2.2 出租车空车率的分类

为了更为直观以及减低随机性的影响,将空车率分为3种水平:高、中、低。分类标准为:将出租车空车率的标准差进行升序排列,视标准差在前50%的司机为稳定司机,删去标准差值处于后50%的司机空车率数值。之后,将空车率的数值从小到大排列,取0~20%、40%~60%以及80%~100%作为低、中、高3种空车率水平的判定标准。对应的空车率总体分布如图2,3种空车率水平的分布如图3所示。观察可得,上海市的出租车空车率大部分分布在中等水平,并且中等空车率水平的标准差最低,最为稳定。

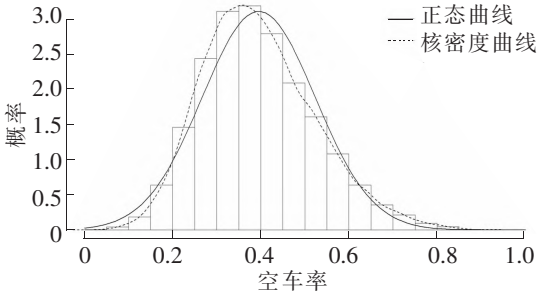

图 2　14:00～16:00 时间段内出租车空车率的分布

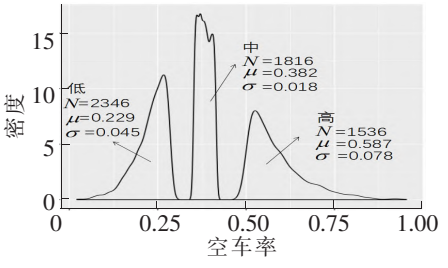

图 3　14:00～16:00 时间段内 3 种空车率水平的分布

2.3　出租车空车率水平影响因素的提取

2.3.1　寻客策略

(1)寻客距离。寻客距离 D_s 是指出租车在上一个乘客的下车事件与紧接着的下一个上车事件之间的空车运行时间,计算如下

$$D_s = \sum_j ED[(lon_j, lat_j),(lon_{j+1}, lat_{j+1})] \tag{2}$$

式中,$ED[(lon_j, lat_j),(lon_{j+1}, lat_{j+1})]$是出租车空车行程中第 j 条记录和它下一条记录之间的欧氏距离,可由经纬度信息计算得到[4]。

(2)上车强度。上车强度 I_p 定义为空车经过沿路一系列网格对应的上车次数的加权平均数,计算如下

$$I_p = \frac{\sum_{(x,y)\in vacanttrip} p^T_{x,y} \times ET^T_{x,y}}{\sum_{(x,y)\in vacanttrip} ET^T_{x,y}} \tag{3}$$

式中,$p^T_{x,y}$为在时间 T 内网格(x,y)中的上车次数;$ET^T_{x,y}$为在网格(x,y)中的第 j 条记录和它上一条记录之间经历的时间。

(3)运行/等待。借鉴 Li et al[5] 的研究,用下列计算来区分司机的运行/等待策略,从而判断司机更倾向于沿路寻客,还是就地等客。定义在上车事件发生前 3 min 的空车运行距离为 D_p,则等待策略对应着指标 D_p 低于一定的阈值 τ_p,而运行策略则对应着指标 D_p 高于该阈值 τ_p,公式表示为

$$D_p \begin{cases} \leqslant \tau_p & \text{等待} \\ > \tau_p & \text{运行} \end{cases} \tag{4}$$

从而建立对应的运行/等待指标 I_w,这是一个布尔变量,当值为 1 时表示司机采用就地等客策略(等待);当值为 0 时,表示司机采用沿路寻客策略(运行),即

$$D_p \begin{cases} \leqslant \tau_p & \text{等待} \quad I_w = 1 \\ > \tau_p & \text{运行} \quad I_w = 0 \end{cases} \tag{5}$$

2.3.2　送客策略

当司机载有乘客时,有的司机偏向于选择保证较高运行速度但较为迂回的道路,有的司机则偏好选择最短路,这些选择最终会通过改变重车时间的占比来影响对应的出租车空车率。

(1)送客迂回程度。送客迂回程度 C_d 通过一次重车行程的实际运行距离和起讫点之间的欧式距离的比值衡量,计算如下

$$C_d = \frac{D_d}{ED[(lon_o, lat_o),(lon_d, lat_d)]} \tag{6}$$

式中,$ED[(lon_o, lat_o),(lon_d, lat_d)]$是起讫点之间的欧氏距离;$D_d$ 为一次重车行程的实际运行距离,由相邻两记录之间的欧氏距离累加得

$$D_d = \sum_j ED[(lon_j, lat_j),(lon_{j+1}, lat_{j+1})] \tag{7}$$

(2)送客速度。为减少 GPS 数据采集间隔的非均质性,送客速度 v_d 为加入时间考虑的速度加权平均值,计算如下

$$v_d = \frac{\sum_j v_j \times ET_j}{\sum_j ET_j} \tag{8}$$

式中,ET_j 为一次重车行程中第 j 条记录和它上一条记录之间经历的时间。

2.4 广义多水平定序 Logit 模型

视 3 种空车率水平为定序的离散因变量 y_i(1 = 高,2 = 中,3 = 低),在前文中提取出的 5 个因素作为自变量 $X_i = (x_{i1}, x_{i2}, \cdots, x_{i5})$,构建 GMOL 模型,那么因变量对应的概率计算如下

$$\begin{cases} p(y_i = 1) = \dfrac{1}{1 + e^{\alpha_1 + X_i^{(1)}\beta^{(1)T} + X_i^{(2)}\beta_1^{(2)T}}} \\ p(y_i = j) = \dfrac{1}{1 + e^{\alpha_j + X_i^{(1)}\beta^{(1)T} + X_i^{(2)}\beta_j^{(2)T}}} - \dfrac{1}{1 + e^{\alpha_{j-1} + X_i^{(1)}\beta^{(1)T} + X_i^{(2)}\beta_{j-1}^{(2)T}} -} \\ p(y_i = M) = 1 - \dfrac{1}{1 + e^{\alpha_{M-1} + X_i^{(1)}\beta^{(1)T} + X_i^{(2)}\beta_{M-1}^{(2)T}}} \end{cases} \tag{9}$$

则广义线性形式的模型可表示为

$$\text{logit}P(y_i \leqslant j) = \ln \frac{P(y_i \leqslant j)}{P(y_i > j)} = \alpha_j + X_i^{(1)}\beta^{(1)T} + X_i^{(2)}\beta_J^{(2)T} \tag{10}$$

式中,$\beta^{(1)}$ 是服从平行线假设的自变量对应的系数,也就是对于任意的空车率水平 j,对应的系数均为 $\beta^{(1)}$;$\beta^{(2)}$ 是违反平行线假设的自变量对应的系数,它们的值随着不同的空车率水平而产生变动。上述系数的值可以通过最小二乘法估计得到。

3 模型结果

3.1 多重共线性检验结果

多重共线性是指多元回归模型中 2 个或 2 个以上独立变量高度相关的现象。方差膨胀因子(*VIF*)是一种检验多重共线性的方法。当 *VIF* 等于 1 时,意味着没有多重共线性存在;当 *VIF* 超过 4 时,则需要进一步进行讨论;而当 *VIF* 超过 10 时,则意味着存在严重的多重共线性问题。对可能影响空车率的因素进行多重共线性检验,结果如表 1 所示。变量对应的 *VIF* 值变动范围为[1,1.84],所有的值均小于 4。因此,可认为提取出的 5 个变量之间不存在明显的多重共线性问题。

表 1 变量的多重共线性结果和平行线假设检验结果

变量	多重共线性检验				平行线假设检验
	VIF	*VIF* 的平方根	容限度	R^2	
寻客距离	1.50	1.22	0.668 4	0.331 6	0.000 0**
上车强度	1.51	1.23	0.664 0	0.336 0	0.000 0**
运行/等待	1.10	1.05	0.906 0	0.094 0	0.000 8**
送客迂回程度	1.00	1.00	0.996 7	0.003 3	0.391 8
送客速度	1.80	1.34	0.554 7	0.445 3	0.995 1

注:** 表示 0.05 水平显著。

3.2 平行线假设检验

平行线假设检验用来分析在不同的空车率水平下,因素对空车率水平造成的影响是否发生改变。由表 1 可知,只有送客迂回程度(C_d)以及送客速度(v_d)在 0.05 水平不显著,服从平行线假设,也就是说这 2 个变量的系数将在不同的空车率水平下分别保持恒定。而其它 3 个变量违反了平行线假设,在不同的空车率水平下,这 3 个变量产生的影响将发生变化。可能的解释是送客迂回程度、送客速度因素对于不同

的空车率水平造成了同等的影响,而其它 3 个变量则会对多样化空车率水平产生显著的影响。因为 GMOL 模型不需要严格遵循平行线假设,故而上述结果亦证明了建立 GMOL 模型的必要性。

3.3 GMOL 模型结果

借助 Stata 14.0 软件中的 gologit2,求得对应的 GMOL 结果如表 2 所示。由于空车率水平有 3 种,因此,$P(y_i \leqslant 3)=1$,表 2 中仅给出了 $P(y_i \leqslant 1)$ 和 $P(y_i \leqslant 2)$ 的结果。送客迁回程度、送客速度的系数在不同空车率水平下恒为 0.115 8,-0.098 8。其它因素的系数的正负性保持一致,说明这些因素对于空车率水平变化方向的影响恒定。

表 2　不同空车率水平下的 GMOL 模型结果

y_i	变量	系数	标准差	$P>\|Z\|$	95%置信区间
≤1	寻客距离***	-1.268 8	0.042 9	0.000	(-1.352 9,-1.184 6)
	上车强度***	0.017 6	0.001 8	0.000	(0.014 0,0.021 1)
	运行/等待***	-0.583 0	0.092 0	0.000	(-0.763 3, -0.402 8)
	送客迁回程度***	0.115 8	0.038 5	0.003	(0.040 3,0.191 2)
	送客速度***	-0.098 8	0.008 0	0.000	(-0.114 6,-0.083 1)
	常数	2.040 2	0.250 1	0.000	(1.549 9,2.530 4)
≤2	寻客距离***	-2.257 2	0.070 8	0.000	(-2.396 0,-2.118 3)
	上车强度***	0.007 6	0.001 6	0.000	(0.004 5,0.010 7)
	运行/等待 **	-0.202 1	0.085 3	0.018	(-0.369 3, -0.035 0)
	送客迁回程度***	0.115 8	0.038 5	0.003	(0.040 3,0.191 2)
	送客速度***	-0.098 8	0.008 0	0.000	(-0.114 6,-0.083 1)
	常数	1.181 6	0.261 4	0.000	(0.669 3,1.694 0)

注:对数似然值 = -3 640.796 1;LRχ^2 = 5 062.41;$P>\chi^2$ = 0.000 0;PseudoR^2 = 0.410 1;观测数 = 5 698。*** 表示 0.01 水平显著,** 表示 0.05 水平显著,* 表示 0.1 水平显著。

对于寻客策略,增加寻客距离将会增加高空车率水平出现的概率,对于那些倾向于在距离上一个乘客下车点更远的地方搜寻下一个乘客的司机而言,他们更容易出现高空车率的情况,应尽量缩短寻客距离。增加上车强度,会减少高空车率水平出现的概率,这表明高空车率水平的司机需要在那些热门区域寻找下一位上车乘客。对于运行/等待而言,选择就地等待策略的司机会拥有更高的空车率水平,那么高空车率司机就需要多采取沿路开车寻客的运行策略来降低他们的空车率。

对于送客策略而言,减少送客的迁回程度、提高送客速度则会增加高空车率水平出现的概率。那么对于那些想提高送客效率的空车率水平较低的司机而言,他们需要采取合理的送客路径规划,或者选择最短路径去送客,或者选择虽然路径更为迁回但能够保证良好的送客速度的道路,比如高架快速路。

结合上述计算结果,对于每一个空车率水平的概率的数学表达形式如下

$$\begin{cases} P(y_i=\text{高})=\dfrac{1}{1+\mathrm{e}^{2.040\,2-1.268\,8D_s+.0176I_p-.5830I_w+.1158C_d-.0988v_d}} \\ P(y_i=\text{中})=\dfrac{1}{1+\mathrm{e}^{1.1816-2.2572D_s+.0076I_p-.2021I_w+.1158C_d-.0988v_d}}-\dfrac{1}{1+\mathrm{e}^{2.040\,2-1.268\,8D_s+.0176I_p-.5830I_w+.1158C_d-.0988v_d}} \\ P(y_i=\text{中})=\dfrac{1}{1+\mathrm{e}^{1.1816-2.2572D_s+.0076I_p-.2021I_w+.1158C_d-.0988v_d}} \end{cases} \tag{11}$$

4　结论

从驾驶员行为角度入手,在司机的送客策略、寻客策略两方面挖掘出了可能影响出租车空车率的 5 个因素,并基于 GMOL 模型,提出一种探究空车率及其影响因素之间关系的量化方法,得到结论如下。

(1)空车率的分布近似于正态分布。空车率的分布显示大部分的上海市出租车空车率分布在中等水平,意味着上海市的总体空车率水平良好。

(2)不同的运行策略会导致不同的空车率水平。高空车率水平的司机偏好远距离寻找乘客、不在热门区域搜寻乘客、倾向于就地等待乘客、或者路径选择不好,选择了又绕又堵的路径送客。

(3)高、低空车率水平的司机需采取不同的运行策略来平衡出租车运行效率和乘客满意度之间的矛盾。对于想降低空车率水平的司机而言,可采取以下策略:缩短寻客距离、采用运行寻客策略、在需求热门区域寻找乘客。而对于想提高送客效率、增加空车率水平的司机而言,则需要通过路径选择来减少重车时间占比。

参 考 文 献

[1]Hu X, Gao S, Chiu Y C, et al. Modeling routing behavior for vacant taxicabs in urban traffic networks[J]. Transportation Research Record: Journal of the Transportation Research Board, 2012, 2284: 81-88.

[2]关金平, 朱竑. 基于FCD的出租车空驶时空特性及成因研究——以深圳国贸CBD为例[J]. 中山大学学报:自然科学版, 2010, 49(s1):29-36.

[3]鞠炜奇, 杨家文, 林雄斌. 城市出租车空载率时空特征及其影响因素研究——以深圳市为例[J]. 规划师, 2015, 31(s2):257-262.

[4]Liu L, Andris C, Ratti C. Uncovering cabdrivers' behavior patterns from their digital traces[J]. Computer Environment & Urban Systems, 2010, 34(6): 541-548.

[5]Li B, Zhang D, Sun L, et al. Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset[C]//Pervasive Computing and Communications Workshops (PERCOM Workshops). [S. l.]:[s. n.], 2011: 63-68.

[6]Moreira-Matias L, Gama J, Ferreira M, et al. Predicting taxi-passenger demand using streaming data[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(3): 1393-1402.

[7]颜研, 倪少权. 基于巢式Logit模型的城市轨道交通客流分配问题研究[J]. 石家庄铁道大学学报:自然科学版, 2015, 28(4):99-103.

[8]Qin G, Li T, Yu B, et al. Mining factors affecting taxi drivers' incomes using GPS trajectories[J]. Transportation Research Part C: Emerging Technologies, 2017, 79: 103-118.

[9]Jun M J, Choi K, Jeong J E, et al. Land use characteristics of subway catchment areas and their influence on subway ridership in seoul[J]. Journal of Transport Geography, 2015, 48: 30-40.

Influencing Factors Analysis of Taxi Vacant Ratio

Tang Juanyu, Zhu Yi, Huang Yizhe

(School of Naval Architecture, Ocean & Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: Taxis are essential means of transport in metropolitans. The taxi vacant ratio level has attracted increased attention from taxi drivers, passengers, and regulators. A comprehensive understanding of the various factors affecting the vacant ratio level is important for the balance between taxi operation efficiency and passenger waiting time. This paper is based on the whole trajectories of taxis, and attempts to find out factors affecting vacant ratio and quantifies the corresponding influences. To differentiate the taxi vacant ratios, the drivers are firstly classified into three vacant ratio levels. By analyzing the taxi vacant trips and occupied trips separately, five factors which may affect the vacant ratio level are extracted from the perspective of passenger searching and delivery strategies. Lastly, a generalized multi-level ordered Logit (GMOL) model is built to identify the significant factors. The results show that these five factors are significant in affecting the taxi vacant ratio, and each factor has different effect on the occurrence of different vacant ratio levels. Therefore, the results of this study consequently provide guidelines for adjusting the vacant ratio level and optimizing the operational management of taxis.

Key words: taxi; vacant ratio; GPS; ordered Logit model; significant factors