

巡游出租车运力规模动态调整回归树模型

叶骥铭¹, 叶晓飞^{1,2*}, 李敏¹, 郑彭军^{1,3,4}, 谢金¹

(1. 宁波大学 海运学院, 浙江 宁波 315832; 2. 宁波市港口贸易合作与发展协同创新中心, 浙江 宁波 315832;

3. 国家道路交通安全工程技术研究中心 宁波大学分中心, 浙江 宁波 315832;

4. 现代城市交通技术江苏高校协同创新中心, 江苏 南京 210096)

摘要: 针对巡游出租车运力规模调整指标体系的匮乏, 以宁波市出租车信息管理系统运营数据为基础, 梳理了巡游车运力规模调整的主要影响因素, 综合考虑各项运营指标, 构建了巡游车运力规模调整阈值回归树模型, 并采用方差分析的卡方自动交叉诊断器算法对模型进行标定。最后, 耦合了万人拥有量、出租车在公交出行结构分担率、巡游车与网约车业务分担比以及乘客平均等候时间等指标在供需状态中的表征作用和重要性排序关系, 提出了巡游出租车运力规模动态调整机制及关键指标的阈值标准, 为城市巡游出租车运力规模调整提供理论依据和决策支持。

关键词: 交通工程; 运力规模; 回归树; 巡游出租车; 阈值分析

中图分类号: U491.1

文献标志码: A

文章编号: 1001-5132 (2020) 04-0089-08

出租车是城市综合交通运输体系的重要组成部分, 但是随着网约车的不断发展, 巡游出租车行业受到了巨大的冲击^[1-2]。国务院办公厅发布的关于深化改革推进出租汽车行业健康发展的指导意见^[3]明确指出通过制定合理的出租车运力规模调整体系来逐步实现市场调节, 这是政府对出租车行业结构实施宏观调整的重要手段, 也是缓解行业矛盾和促进出租车行业良性发展的重要保障。

国内外学者对出租车运力规模预测及监测评估展开了深入研究^[4-9], 指出出租车运力规模调整关键指标及阈值包含: 万人拥有量(介于 20~30 辆·万人⁻¹, 大城市每万人不少于 20 辆, 小城市不少于 5 辆, 中等城市取中间值^[10])、出租车占公共交通分担比(介于 10%~20%)、里程利用率或空驶率(阈值为 60%^[4]、65%^[5]、70%^[11])、乘客平均等车时间(浙江省平峰 10 min、高峰期 15 min^[4]、宁波市期望值为 6.72 min, 实际值为 12.49 min^[4]、大连市期望值为 3.75 min^[5])、百公里营收或单车日均营收(与成本支出、业务量等多种因素相关, 无明确的定

量标准)、道路交通拥堵指数或单差等候时间(无明确阈值标准)、单车日均车次(无阈值)等指标。已有研究表明, 巡游出租车运力规模调整指标并无统一体系, 各个关键指标阈值无明确的定量标准且缺乏指标间的重要性排序关系, 同时未考虑到与网约车的竞争关系。因此, 本文以出租车信息管理系统运营数据为基础, 综合考虑单车日均车次、里程利用率等指标, 采用数据挖掘-决策树方法构建巡游车运力规模调整回归树模型, 提出巡游车动态调整机制及关键指标的阈值和排序关系, 为城市巡游车运力规模调整提供理论依据和决策支持。

1 巡游车运力规模影响因素分析

巡游出租车运力规模影响因素众多, 宏观层面包括城市性质和职能、旅游吸引能力、社会经济发展、城市规模和布局、自然地理条件、城市交通基础设施和公交服务质量; 微观层面有万人拥有量、出租车占公共交通分担比、里程利用率、单车日均营收等指标, 这是描述巡游车供需匹配程度

收稿日期: 2019-12-16.

宁波大学学报(理工版)网址: <http://journal.lg.nbu.edu.cn/>

基金项目: 浙江省自然科学基金(LY20E080011); 国家自然科学基金(71971059); 国家重点研发计划(2017YFE9134700); 宁波市交通运输科技项目(201920)。

第一作者: 叶骥铭(1996-), 男, 浙江宁波人, 在读硕士研究生, 主要研究方向: 交通运输规划与管理。E-mail: 1115635697@qq.com

*通信作者: 叶晓飞(1985-), 男, 浙江宁波人, 博士/副教授, 主要研究方向: 交通运输规划与管理。E-mail: yexiaofei@nbu.edu.cn

的关键指标。本节以宁波市巡游车运营数据为基础,重点探讨网约车对巡游车运营效率的影响。

图1所示为从2014至2017年在不同滴滴营销活动和关键业务推出的情况下巡游车月均总车次的变化。从2014年1月至3月,滴滴依靠巡游车抢占市场,市场占有率达60.2%;随着专车业务的推出,巡游车靠着接入滴滴等平台提高供需匹配度,营运业务量保持高水平;然而,伴随着滴滴和快的合并以及快车、顺风车和拼车业务的推出,大批私人小汽车加入滴滴等网约车平台(神州专车等平台是自有车辆)涌入客运市场,致使巡游出租车业务量水平显著下降;直至网约车新政颁布以及滴滴和优步合并及其提价方案的实施,加上网约车的合法化,巡游车业务量水平下降趋势得以抑制,其业务量总体亦趋于稳定。

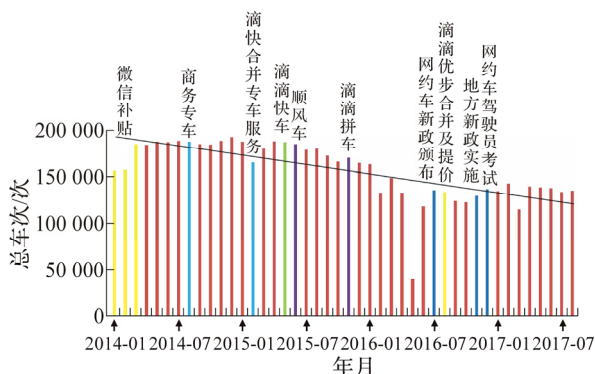


图1 巡游出租车月均总车次与网约车营销活动对应关系

与2014年相比,2017年巡游出租车各项运营指标显著下降(图2),日均里程利用率由68.70%下降为58.55%,降幅为14.78%;单车日均车次由43.10趟次下降为31.29趟次,降幅为27.40%;百公里营收水平由234.41元·百公里⁻¹下降为223.98元·百公里⁻¹,降幅为4.45%;单车日均营收水平由939.47元·车⁻¹下降为722.82元·车⁻¹,降幅为23.06%;日均单车行驶里程由398 km下降到327 km,降幅为17.84%。2016年8月以来,巡游车运营效率虽有所下降,但基本上趋于稳定。可以看出网约车与巡游车业务量水平是此消彼长的关系,因此,巡游车运力规模调整时必须要考虑网约车的影响,拟用二者业务分担比指标来描述其影响。综上,本文选取的巡游车运力规模影响因素有里程利用率、百公里营收、单车日均车次、单车日均营收、单差平均等候时长和单车日均营运时长。

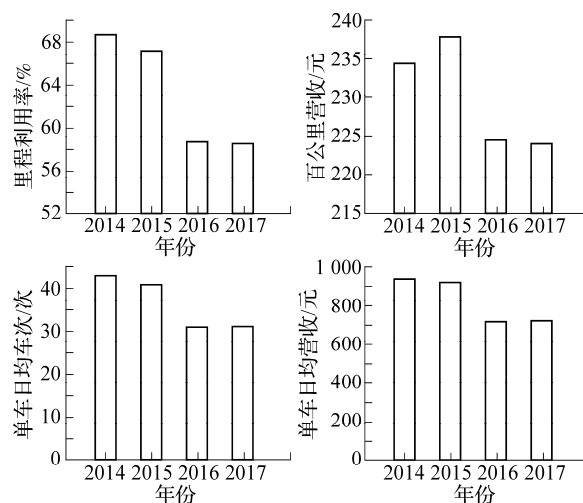


图2 巡游车运营指标变化

2 回归树模型及算法

2.1 回归树模型

表征巡游车运力充足与否的影响指标众多,调整其运力规模不适宜用单一标准,且权重体系评价方法难以给出各项指标间的明确排序及耦合关系^[12]。决策树作为一种实现分治策略的层次数据结构,采用优先策略自上而下进行不断细分和输出预测结果,并可转换成易懂的简单规则^[13],其中,采用了回归的不纯性度为划分标准。

假设某一关键指标如里程利用率对应决策节点 m ,令 X_m 为 X 到达节点 m 的子集,即它是 $x \in X$ 的满足从树根到节点 m 的所有决策点条件的所有 x ,则决策节点分类结果为

$$b_m(x) = \begin{cases} 1, & \text{如果 } x \in X_m : x \text{ 到达节点 } m, \\ 0, & \text{否则.} \end{cases} \quad (1)$$

以估计值的均方误差 E_m 为划分度量标准,令 g_m 为节点 m 中的估计值,

$$E_m = \frac{1}{N_m} \sum_t (r^t - g_m)^2 b_m(x^t), \quad (2)$$

$$\text{式中: } N_m = |X_m| = \sum_t b_m(x^t); \quad g_m = \frac{\sum_t b_m(x^t) r^t}{\sum_t b_m(x^t)}; \quad t$$

为序号; r 为输入值; r^t 为序号为 t 的输入值。

E_m 对应于 m 上的方差。如果误差是可以接受的, $E_m < \theta_r$,则创建一个树节点,存放 g_m ;否则,达到节点 m 的数据进一步划分,使得分支的误差和最小。在每个节点上,寻找最小化误差的属性和

数值属性的划分阈值,然后递归地进行上述过程。

令 X_{mj} 为 X_m 的分支 j 子集: $\bigcup_{j=1}^m X_{mj} = X_m$, 定义

$$b_{mj}(x) = \begin{cases} 1, & \text{如果 } x \in X_{mj}: x \text{ 到达节点 } m \\ & \text{并取分支 } j, \\ 0, & \text{否则.} \end{cases} \quad (3)$$

令 g_{mj} 是到达节点 m 的分支 j 的估计值,

$$g_{mj} = \sum_t b_{mj}(x^t) r^t / \sum_t b_{mj}(x^t). \quad (4)$$

而划分后的误差为

$$E'_m = \frac{1}{N_m} \sum_j \sum_t (r^t - g_{mj})^2 b_{mj}(x^t). \quad (5)$$

对于任意划分,误差的减少由式(2)和式(5)之差给出。寻找误差减少最大化的划分方向或等价于式(5)取最小值。均方误差是一种可能的误差函数,另一种最大误差

$$E_m = \max_j \max_t |r^t - g_{mj}| b_{mj}(x^t). \quad (6)$$

应用最大误差可保证任意实例的误差都不大于给定阈值。可接受的误差阈值是复杂度的函数,其值越小,产生的树越大且过分拟合的风险越大;其值越大,拟合不足和过分光滑的可能性越大。

当每个决策节点使用所有关键性指标作为输入维时,则线性多变量节点定义为

$$f_m(x): \mathbf{w}_m^T x + \mathbf{w}_{m0} > 0. \quad (7)$$

从根到树叶的路径上的相继节点进一步划分实例,而叶节点定义输入空间上的多面体,逐一对权重 w_{mj} 进行微调来降低统计显著性指标,通过子集选择降低维度并降低节点的复杂度。

2.2 模型算法

出租车管理信息系统中因变量和解释变量均为数值型,针对这一数据结构特点,采用方差分析的卡方自动交叉诊断器算法作为构建回归树的算法。对于经过预处理的输入变量,计算与输出变量相关性检验的统计量的概率 $-p$ 值。对于数值型输出变量,采用 F 统计量。概率 $-p$ 值最小的输入变量与输出变量的联系最紧密,应作为当前最佳分组变量。当 $-p$ 值相同时,应选择检验统计量观测值最大的输入变量,不再重新确定分割点,而自动将分组变量的各个类别作为树分支,形成多叉树。

回归树的模型风险评价指标主要是预测值的

方差和误判率,其中风险轨迹的标准误差为

$$\alpha_w = \sqrt{p(1-p)/N}, \quad (8)$$

式中: p 为风险估计; N 为样本量。

利用保持方法进行回归树模型性能评估,将被标记的原始数据划分为两个不相交的集合,分别为训练集和检验集。在训练数据集上归纳模型,在检验集上评估模型的性能。

3 模型标定及结果分析

3.1 建模数据来源及分析

采用2014年1月至2017年8月的宁波市出租车运营数据进行建模分析,数据来源于宁波市客运管理局出租车运营管理信息系统,包含有效样本数据1317组,其中训练样本856组,约占65%,用于决策树构造,其余35%用于模型检验。从2001至2015年,营运权通过服务质量招投标形式配置给企业的车辆数共计1716辆。通过对宁波市出租车运营数据的分析可知宁波市出租汽车拥有量仍有待发展,然而,2013至2017年巡游车保有量保持在4427辆,未做调整。

宁波市巡游车运营关键指标分布情况见表1,其中运力规模作为回归树的因变量,是根据供需平衡法计算获得的^[14],略高于实际运营数量。总车次、电召次数、总营收、行驶里程、营运里程、营运时间、总等候时长是总量指标,与里程利用率、单车日均车次、单车日均营收等指标有共线性问题,并且后者更能表现出单车运营效率,故选用后者作为回归树模型的解释变量。此外,电召服务系统是从2014年6月开始投入使用,电召次数占总车次比重较低,空驶率与里程利用率之和为1,选用里程利用率作为代表指标。

3.2 决策树结构及结果分析

运用回归树方法对营运数据进行分类,以运力规模预测值为目标属性,对样本进行有监督的分类,模型置信度为87.36%,说明较好地归纳和解释了巡游车运力规模调整阈值。如图3所示,该回归树是深度为3的多叉树型结构,根节点为运力规模预测值,矩形表示叶节点,分别给出了每个节点的决策变量、样本数、样本均值与预测均值、 F 统计值以及概率 $-p$ 值。里程利用率、单车日均营收等影响因素按相关性递减依次纳入到模型中,

表1 宁波市巡游出租车营运指标描述性统计

数值类型	实际运营 车辆数/辆	总车次/次	电召次数/ 次	总营收/元	行驶里程/ km	营运里程/ km	营运时间/ min	总等候 时长/s
均值	4 211	172 577	1 413	3 848 260	1 630 604	1 097 277	2 549 508	9 792.48
极小值	1 103	31 714	0	745 433	333 833	193 486	459 440	2 090.94
极大值	4 387	226 861	15 056	4 876 810	2 023 800	1 431 900	4 152 456	15 729.37
85%位数	4 362	191 954	1 649	4 250 160	1 767 389	1 234 745	2 879 151	12 679.16
数值类型	里程 利用率	百公里 营收/元	单车日均 车次	单车日均 营收/元	单差平均 等候时长/s	单车日均营 运时长/min	运力规模 预测值/辆	
均值	0.669 9	235.53	40.80	910.67	209.49	602.85	4 841	
极小值	0.567 6	200.64	20.26	522.03	127.55	310.42	1 268	
极大值	0.750 3	262.20	55.57	1 186.98	433.02	970.20	5 043	
85%位数	0.700 4	243.88	44.66	988.71	254.93	670.99	5 014	

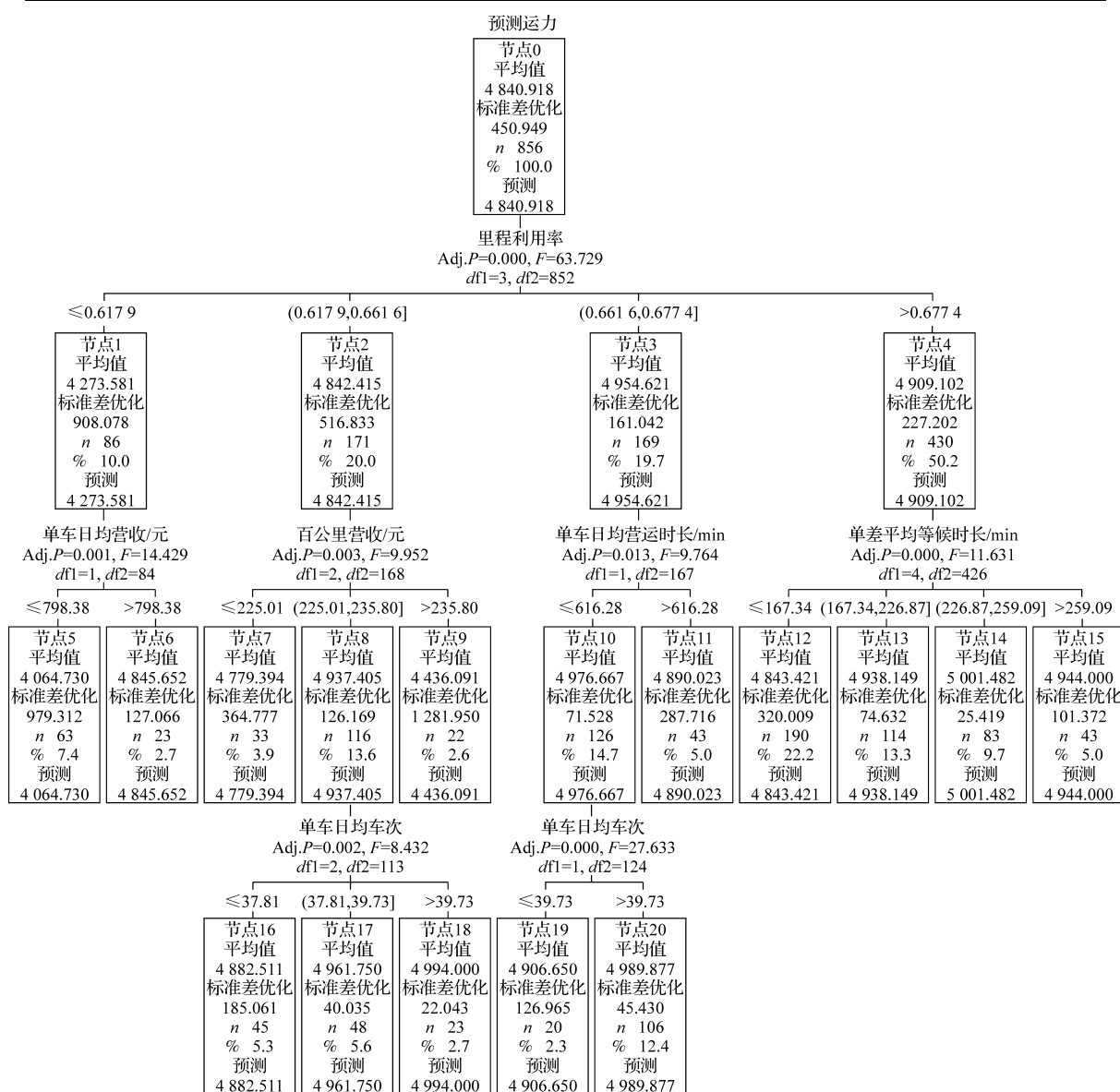


图3 巡游车运力规模回归树结构

说明对运力规模影响显著。其中,里程利用率影响程度最高, F 统计值为63.73;其次为单车日均营收、单差平均等候时长、单车日均营运时长和百公里营收, F 统计值依次递减,说明影响程度递减;最后,单车日均车次被选为第3层分类规则指标,且各层指标概率 $-p$ 值均小于0.05,均通过显著性检验。

根据决策树分析结果,生成了14条规则,用“IF-THEN”的形式表示(表2),结合巡游车运价与运力相互作用关系,归纳总结巡游车运力规模调整与投放特征如下:

节点5,里程利用率偏低,表明巡游车供过于求,且运营效率较低;单车日均营收偏低,说明驾驶员收入偏低。此阈值条件下,实际运力规模过剩,需缩减运力($4427-4065=362$ 辆),以达到供需平衡。

节点6,里程利用率偏低,但单车日均营收水平说明驾驶员收益尚可,此阈值条件下,运力规模无需调整,但为了提高里程利用率,可通过价格杠杆(降价)来提高客流吸引率。

节点7,里程利用率处于(0.6179, 0.6616]区间内,百公里营收 ≤ 225.01 ,略小于实际运价基本费率值($2.35 \text{ 元}\cdot\text{km}^{-1}$),表明巡游车经营状况不佳,但该决策节点运力预测值略低于父节点2,运力规模调整窗口不打开,应适应该阈值条件下运价基本费率值以及以提高营收水平为目标,通过降低运价方式适应此供需状态下的市场状况。

节点16,里程利用率和百公里营收均处于可接受范围,但单车日均车次偏低,说明客运需求不足,因此,运力规模调整窗口未打开,可通过降价方式来提高客流需求。

表2 回归树算法生成提取的主要类别

决策节点	IF	THEN	F 统计值	归类
5	(里程利用率 ≤ 0.6179)AND(单车日均营收 ≤ 798.38)	4 065	14.43	运力过剩,压缩运力362辆
6	(里程利用率 ≤ 0.6179)AND(单车日均营收 > 798.38)	4 845	14.43	运力调整窗口未打开,但需降价吸引客流
7	($0.6179 < \text{里程利用率} \leq 0.6616$)AND(百公里营收 ≤ 225.01)	4 780	9.95	运力调整窗口未打开,但需降价适应低价市场
16	($0.6179 < \text{里程利用率} \leq 0.6616$)AND($225.01 < \text{百公里营收} \leq 235.8$)AND(单车日均车次 ≤ 37.81)	4 883	8.43	运力调整未打开,但需降价吸引客流
17	($0.6179 < \text{里程利用率} \leq 0.6616$)AND($225.01 < \text{百公里营收} \leq 235.8$)AND($37.81 < \text{单车日均车次} \leq 39.73$)	4 962	8.43	运力调整窗口未打开,无需调整运力
18	($0.6179 < \text{里程利用率} \leq 0.6616$)AND($225.01 < \text{百公里营收} \leq 235.8$)AND(单车日均车次 > 39.73)	4 994	8.43	运力调整窗口未打开,但需加价抑制客流
9	$0.6179 < \text{里程利用率} \leq 0.6616$)AND(百公里营收 > 235.8)	4 437	9.95	运力调整窗口未打开,但需加价适应高价市场
19	($0.6616 < \text{里程利用率} \leq 0.6774$)AND(单车日均营运时长 ≤ 616.28)AND(单车日均车次 ≤ 39.73)	4 907	27.63	运力调整窗口未打开
20	($0.6616 < \text{里程利用率} \leq 0.6774$)AND(单车日均营运时长 ≤ 616.28)AND(单车日均车次 > 39.73)	4 990	27.63	运力调整窗口未打开,但需加价抑制客流
11	($0.6616 < \text{里程利用率} \leq 0.6774$)AND(单车日均营运时长 > 616.28)	4 890	9.76	调整窗口打开,投放运力463辆
12	(里程利用率 ≥ 0.6774)AND(单差平均等候时长 ≤ 167.34)	4 715	11.62	运力调整窗口未打开
13	(里程利用率 ≥ 0.6774)AND($167.34 < \text{单差平均等候时长} \leq 226.87$)	4 807	11.62	运力调整窗口未打开,加收等候费用
14	(里程利用率 ≥ 0.6774)AND($226.87 < \text{单差平均等候时长} \leq 259.09$)	4 869	11.62	运力调整窗口未打开,提高等候费率标准且发布调整预警
15	(里程利用率 ≥ 0.6774)AND(单差平均等候时长 > 259.09)	4 812	11.62	调整窗口打开,投放运力385辆

节点 17, 各项指标均处于可接受范围, 运力规模调整窗口未打开, 无需调整运价。

节点 18, 与节点 16 相反, 单车日均车次偏高, 说明客运需求旺盛, 运力规模调整窗口未打开, 可通过加价方式来抑制客运需求。

节点 9, 与节点 7 相反, 百公里营收 ≥ 235.8 , 略高于实际运价基本费率值($2.35 \text{ 元}\cdot\text{km}^{-1}$), 运力调整窗口未打开, 通过加价适应高价市场状况。

节点 19, 里程利用率处于 $(0.6616, 0.6774]$ 区间内, 该阈值偏高, 说明需求略大于供给, 但是单车日均运营时长 ≤ 616.28 , 说明驾驶员工作强度未超过极限忍耐值; 以及单车日均车次 ≤ 39.73 , 未超过极限值, 说明需求与供给之差尚可接受, 无需进行运力规模调整。

节点 20, 与节点 19 相反, 单车日均车次超过极限值, 说明需求与供给之差较大, 可适当加价激励驾驶员延长一定工作时间以降低里程利用率、单车日均车次和抑制客运需求。

节点 11, 与节点 19 和 20 类似, 里程利用率偏高, 但单车日均运营时长超过极限值, 驾驶员工作强度太高, 不利于行业良性发展。因此, 调整窗口打开, 投放运力($4890-4427=463$ 辆)。

节点 12, 里程利用率大于 0.6774 , 该阈值处于较高水平, 说明需求大于供给, 但单差平均等候时长 ≤ 167.34 , 说明城市交通拥堵状况影响较小, 巡游车在高里程利用率条件下服务质量仍处于较高水平, 不建议调整运力规模。

节点 13, 单差平均等候时长加大, 交通拥堵影响较大, 巡游车服务等候时间增加, 为弥补出租车低速行驶成本, 建议在运价标准当中加收等候费。

节点 14, 单差平均等候时长进一步加大, 交通拥堵状况恶化, 出租车在高里程利用率条件下完成服务需要付出额外成本, 建议提高等候费率来弥补额外成本支出。

节点 15, 单差平均等候时长超过忍耐值, 说明城市交通状况剧烈恶化, 出租车在高里程利用率条件下完成服务需要付出高额成本, 为了降低里程利用率和弥补低速行驶成本, 达成新的供需平衡, 运力规模调整窗口打开, 投放适量运力($4812-4427=385$ 辆)。

此外, 图 4 分别给出了训练样本与检验样本的

累计收益, 前期快速达到了较高的累计收益, 然后很快趋于 100% 并平稳, 说明回归树模型较为理想。采用预测值的方差和误判率进行回归树模型风险评价, 训练与检验样本标准误差为 6.13%, 误判率为 0.07, 总体上模型的精度较好。

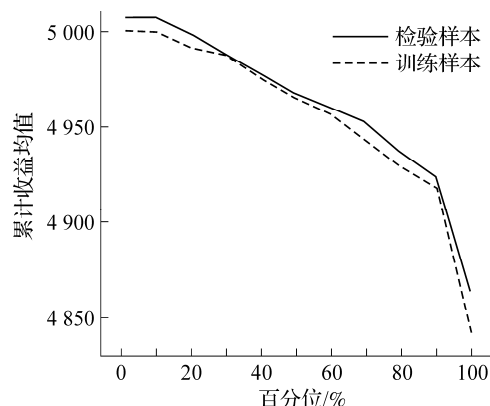


图 4 训练与检验样本累计收益

4 巡游出租车运力规模动态调整机制

总结归纳国内外文献发现, 除了回归树模型中的运营指标之外, 运力调整还需依据以下指标:

(1)万人拥有量, 直接反映了巡游车运力规模与城市人口规模的关系。依据国内外各大城市的经验, 城市出租车万人拥有量介于 20 到 30 辆之间, 能表现出较好的市场接受度。图 5 所示为宁波市与国内其他城市万人拥有量分布, 结合宁波发展情况的多方面因素, 与同等级别城市相比较, 建议近期宁波市巡游车运力规模调整指标——万人拥有量阈值为 $23 \text{ 辆}\cdot\text{万人}^{-1}$ 。

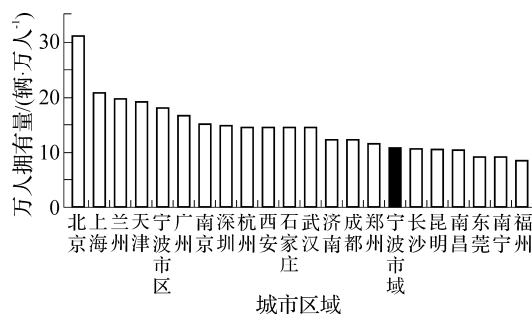


图 5 国内部分城市出租汽车万人拥有量

(2)出租车在公交出行结构分担率。巡游车作为城市公共交通系统的有效补充, 其最优分担率为 8%~20%, 过高阻碍“公交都市”战略和行业的合理发展, 过低则有可能被淘汰。结合 2015 年宁波市 OD 调查结果和公共交通快速发展状况, 建议该

指标阈值为 18%。

(3)巡游车与网约车业务分担比,反映了网约车对巡游车运营效率的冲击与挑战。如图 6 所示,网约车与巡游车业务量对比情况表明,从 2016 年 12 月以来,随着合法化进程加快,巡游车与网约车业务量出现平衡状态,二者分担比在均值 1.53 (10:6.54)左右浮动,且浮动幅度不大。因此,以 10:6.54 作为巡游车与网约车业务分担比的阈值。

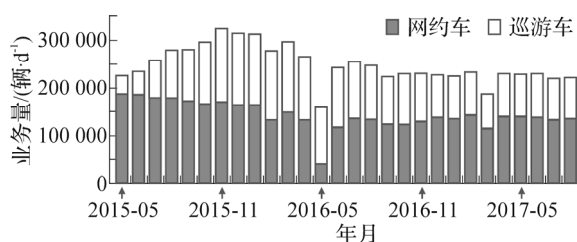


图6 网约车与巡游车业务量对比分析

(4)乘客平均等待时间,反映了出租车的服务质量以及供需状态。采用随机拦截出租车抽样调查了扬召乘客和网约电召乘客的等待时间,排除电话叫车乘客等待时间(以预约业务为主,等待时间较长),分别选取不同时段(高峰与平峰)、不同区域(大专院校、大型商圈、大型公园、旅游景点、交通枢纽、地铁换乘站、大型医院等;随机抽取地点:除以上区域以外的地点)下的时空条件进行调查,结果如图 7 所示,高峰和平峰下扬召和网召的乘客平均等待时间为 13.53、11.72、8.23、5.87 min,建议根据扬召与网召业务量稳定比确定高峰和平峰乘客平均等待时间阈值,分别为 12.82 和 9.61 min。由于乘客平均等候时间单一指标无法准确、客观地评估巡游车供需关系,综合考虑其他指标高峰与平峰的供需特点,建议按照高峰需求弹性(1.1)和平峰供给出车率(0.9)两个因素计算乘客平均等待时间阈值,取为 11.38 min。

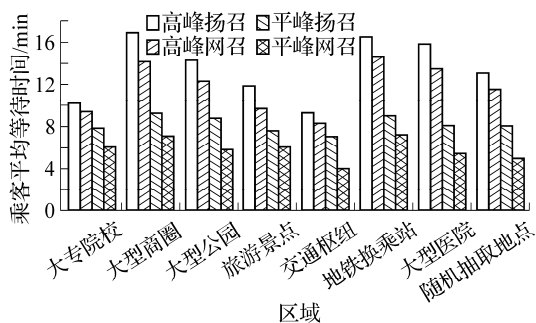


图7 不同时段各地点乘客平均等待时间分布

综上所述,采用单一指标作为评估巡游车运力规模都具有片面性和不准确性,需耦合各个指标共同使用。因此,采用专家打分法,从宏观到微观对指标重要性进行排序。如图 8 所示,给出了调整指标排序及阈值。此外,城市出租车供需平衡是动态变化,可通过监测-调整决策-再监测-再调整决策的不断调节来实现出租车市场需求与运力供给的动态平衡。

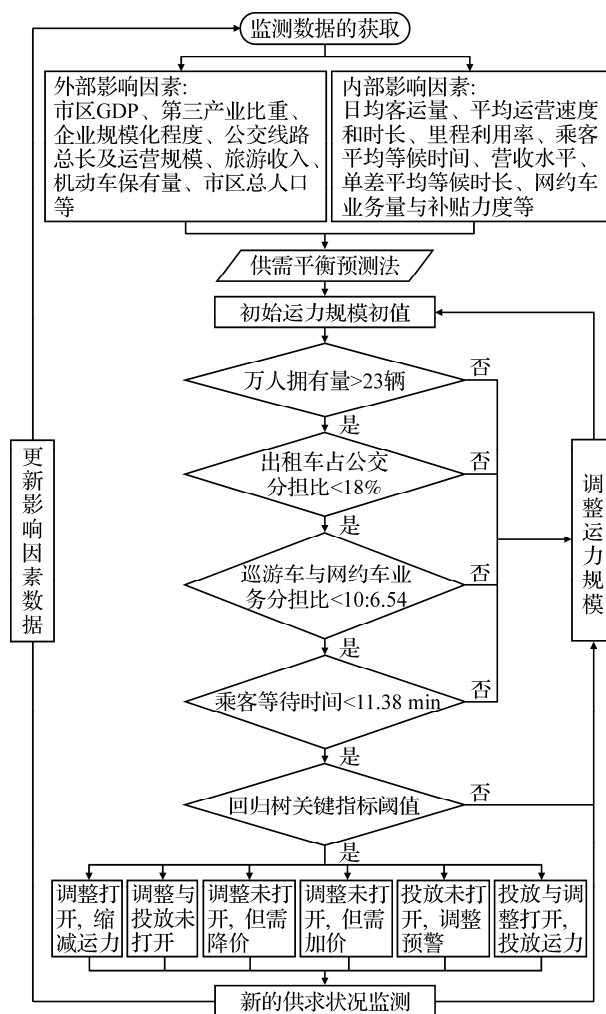


图8 巡游出租车运力规模动态调整机制

5 结论

针对巡游出租车运力规模调整指标体系的匮乏,本文以出租车信息系统运营数据为基础,综合考虑了单车日均车次、里程利用率、营收水平等指标,构建了巡游车运力规模调整回归树模型,得出相应阈值区间,并耦合万人拥有量、出租车在公交出行结构分担率、巡游车与网约车业务分担比以及

乘客平均等候时间等指标在供需状态中的表征作用和重要性排序关系,提出了巡游车动态调整机制及关键指标的阈值标准,结果表明决策树方法和动态调整机制可行有效,同时可为我国其他城市的巡游出租车运力规模调整提供理论依据和决策支持。

参考文献:

- [1] 沈尘. “互联网+出行”模式对传统出租车行业的影响[J]. 对外经贸, 2015, 255(9):76-77.
- [2] 肖赟, 魏朗. “互联网+”背景下我国出租车运价体系改革路径研究[J]. 价格理论与实践, 2016, 23(7):79-81.
- [3] 国务院办公厅. 国务院办公厅关于深化改革推进出租汽车行业健康发展的指导意见[EB/OL]. [2016-07-28]. http://www.gov.cn/zhengce/content/2016-07/28/content_5095567.htm.
- [4] 杨仁法. 宁波市客运出租车行业监控体系研究[D]. 西安: 长安大学, 2007.
- [5] 刘鸿婷. 出租车运力规模评价与优化研究[D]. 大连: 大连海事大学, 2011.
- [6] 栾丽娜. 基于 GPS 数据的城市出租车运营分析与数据挖掘[D]. 济南: 山东大学, 2015.
- [7] Douglas G W. Price regulation and optimal service standards: The taxicab industry[J]. Journal of Transport Economics and Policy, 1972, 6(8):116-127.
- [8] Yang H, Ye M, Tang W H, et al. Regulating taxi services in the presence of congestion externality[J]. Transportation Research Part A: Policy and Practice, 2005, 39(1):17-40.
- [9] Walter S, Juan A M. An international taxi cab evaluation [J]. Knowledge and Process Management, 2010, 34(3): 112-127.
- [10] GB 50220-1995. 城市道路交通规划设计规范[S].
- [11] 曹祎, 罗霞. 打车软件背景下出租车运营平衡模型[J]. 长安大学学报(自然科学版), 2015, 35(增 1):203-207.
- [12] Wong K I, Wong S C, Yang H. Modeling urban taxi services in congested road networks with elastic demand [J]. Transportation Research Part B: Methodological, 2001, 35(9):819-842.
- [13] 薛薇, 陈欢歌. Clementine 数据挖掘方法及应用[M]. 北京: 电子工业出版社, 2010:149-162.
- [14] 陆建, 王伟. 城市出租车拥有量确定方法[J]. 交通运输工程学报, 2004, 4(1):92-95.

Regression tree model for dynamic adjustment of the scale of cruising taxicab capacity

YE Qiming¹, YE Xiaofei^{1,2*}, LI Min¹, ZHENG Pengjun^{1,3,4}, XIE Jin¹

(1.Faculty of Maritime and Transportation, Ningbo University, Ningbo 315832, China;

2.Ningbo Port Trade Cooperation and Development Collaborative Innovation Center, Ningbo 315832, China;

3.Ningbo University Sub-center, National Traffic Management Engineering & Technology Research Center, Ningbo 315832, China;

4.Jiangsu Province Collaborative Innovation Center for Modern Urban Traffic Technologies, Nanjing 210096, China)

Abstract: In view of the lack of indicator system for the scale adjustment of the cruising taxicab capacity, this paper collects the operation datasets of taxicab from the Taxi Information Management System in Ningbo and analyzes the influential factors on cruising taxicab capacity. The regression tree model for the dynamic adjustment of the scale of cruising taxicab capacity is then established by considering the various operational indicators of the cruise taxicab. Next, the chi-squared automatic interaction detector is applied to calibrate the model. Finally, the dynamic adjustment mechanism of the scale of cruising taxicab capacity and the thresholds of the key indicators are proposed by coupling the representation functions on the balance of taxi supply and demand with the ordering relationship of importance of other indicators, which consist of the taxi ownership per person, the sharing ratio of taxi in public transport trip structure, the sharing ratio between cruising taxi and ride-hailing taxi, and the average waiting time of passengers, to provide theoretical basis and decision-making support for the scale adjustment of urban cruise taxi capacity.

Key words: traffic engineering; the scale of capacity; regression tree; cruising taxicab; threshold analysis

(责任编辑 韩 超)