

文章编号: 1009-6744(2020)04-0173-05

中图分类号: U491

文献标志码: A

DOI:10.16097/j.cnki.1009-6744.2020.04.025

基于机器学习模型的建成环境对小汽车拥有行为的影响

王晓全, 邵春福*, 管岭, 尹超英

(北京交通大学 综合交通运输大数据应用技术交通运输行业重点实验室, 北京 100044)

摘要: 为分析家庭小汽车拥有行为,同时考虑居住地和工作地建成环境的影响,构建梯度提升迭代决策树(GBDT)模型;分析社会经济属性,居住地、工作地建成环境属性对小汽车拥有行为的影响程度,并基于长春市居民出行调查数据进行实证研究.结果表明:3类影响因素中,社会经济属性对小汽车拥有行为的影响最大(58.95%);职住地建成环境属性均对家庭小汽车拥有行为具有显著影响,且居住地建成环境影响(23.77%)高于工作地建成环境(17.28%);职住地建成环境属性中,除居住地交叉口密度,工作地到中央商务区(CBD)距离及公共交通站点密度外,其他建成环境属性对小汽车拥有行为的影响均大于5%.因此,有必要同时优化职住地的建成环境来抑制小汽车拥有量的增长.

关键词: 交通工程;职住地建成环境;小汽车拥有;影响程度;梯度提升迭代决策树

Exploring Influences of Built Environment on Car Ownership Based on a Machine Learning Method

WANG Xiao-quan, SHAO Chun-fu, GUAN Ling, YIN Chao-ying

(Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, China)

Abstract: To analyze the car ownership behaviors, a gradient boosting decision tree (GBDT) method is employed to explore the effect sizes of residential and workplace built environments on car-ownership decisions. The empirical analysis is conducted based on the Changchun household travel survey data. The results show that the socio-economic factors contribute 58.95% to automobile ownership collectively and rank the first among the three categories of factors. The residential and workplace built environment variables are both associated with car ownership. And the residential built environment is more influential than the workplace built environment. Except for intersection density at residential locations, distance to the central business district(CBD), and bus stop density at workplace locations, all built environment variables have relative importance more than 5%. Therefore, it is of great importance for urban planners and policy makers to optimize the urban built environment to mitigate the increase of car ownership.

Keywords: traffic engineering; workplace built environment; car ownership; relative importance; gradient boosting decision tree

0 引言

随着城市化进程加快和社会经济水平提高,城市居民对小汽车依赖程度逐渐提高,小汽车依

赖性增强带来的交通拥堵和尾气排放问题给城市交通及环境系统带来了严峻挑战.探究土地利用和小汽车依赖之间的关系被认为是解决小汽车增

收稿日期:2020-02-14

修回日期:2020-05-13

录用日期:2020-05-15

基金项目:中央高校基本科研业务经费专项资金/The Fundamental Research Funds for the Central Universities (2019YJS101).

作者简介:王晓全(1992-),男,黑龙江宾县人,博士生.

*通信作者:cfshao@bjtu.edu.cn

长带来交通及环境问题的重要手段。

既有研究对建成环境与小汽车拥有之间的关系进行了探讨。尹超英等^[1]构建路径分析—离散选择一体化模型分析建成环境对小汽车拥有和使用行为的影响,结果表明,居住地建成环境对小汽车拥有行为具有显著影响。王晓全等^[2]考虑交通小区之间的相关性,基于 Bayesian 层次模型分析了居住地建成环境对小汽车拥有行为的影响,结果表明,控制交通小区间的相关性和个体社会经济属性后,居住地建成环境仍对家庭小汽车拥有行为具有显著影响。丁川^[3]构建考虑空间异质性的多层次序 Probit 模型,研究居住地建成环境对小汽车拥有的影响,结果表明,居住地建成环境对小汽车拥有具有显著影响。Yin 等^[4]同时分析城市层和社区层建成环境对小汽车拥有行为的影响,揭示不同尺度建成环境对家庭小汽车拥有行为的影响。Guo^[5]以纽约市城区调查数据为例,分析居住地建成环境和停车可用性对家庭小汽车拥有行为的影响,发现在控制社会经济属性后,居住地的建成环境属性和停车供给均会影响家庭小汽车拥有行为。既有研究多关注居住地建成环境对小汽车拥有的影响,较少关注工作地建成环境对小汽车拥有行为的影响^[6]。此外,现有研究多是基于传统的统计模型分析建成环境对

小汽车拥有行为的影响,模型结果仅给出建成环境变量和小汽车拥有行为之间的统计显著性,无法量化各个属性对小汽车拥有行为的影响程度,然而各个属性的作用程度对城市规划及交通政策的制定具有重要意义^[7]。因此,有必要构建机器学习模型量化捕捉居住地和工作地建成环境属性对小汽车拥有行为的影响程度。

在既有研究的基础上,本文将居住地和工作地建成环境同时纳入模型,通过构建梯度提升迭代决策树(Gradient Boosting Decision Tree,GBDT)模型量化捕捉建成环境各要素对家庭小汽车拥有的影响程度。研究结果可为通过优化城市土地利用缓解小汽车依赖提供理论支撑。

1 研究数据

以2012年长春市居民出行调查数据为基础,研究职住地建成环境属性对小汽车拥有行为的影响。长春市居民出行调查数据覆盖长春市5个行政区,237个交通小区。剔除缺失数据,共得到5 201条数据用于实证分析,样本数据中居民职住地理位置分布如图1所示。家庭社会经济属性直接从出行调查数据中提取,包括性别、年龄、教育、家庭规模和家庭收入。

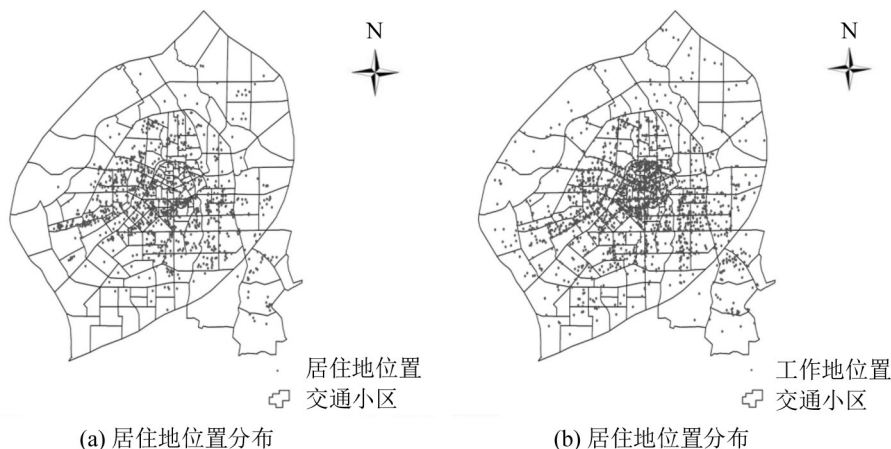


图1 样本职住地理位置分布

Fig. 1 Distribution of residences and workplaces

职住地建成环境属性均包括:到中央商务区(Central Business District, CBD)距离、土地利用混合度、公共交通站点密度及交叉口密度4个变量,分别用于度量建成环境的“目的地可达性”“多样性”“公共交通可达性”“设计”4个维度。建成环境

各变量均基于 ArcGIS 平台计算得到。首先,根据出行调查数据中记录的居民居住地和工作地的地理位置信息对居民的职住地进行定位;其次,根据定位计算职住地到 CBD 的欧几里得距离;同时,根据长春市交通地图计算得到各交通小区的交叉口密

度;最后,基于高德地图 API 提取兴趣点(Point of Interest, POI)数据,计算交通小区公共交通站点密度及土地利用混合度.土地利用混合度基于熵指数法计算得到,计算所利用的 11 种兴趣点数据包括:住宅、宾馆、餐饮、超市、公园、广场、购物中心、学校、医院、银行及政府^[8],计算过程为

$$E = \frac{-\sum p_{ij} \ln p_{ij}}{\ln N_j} \quad (1)$$

式中: p_{ij} 为交通小区 j 内第 i 种兴趣点所占的比例; N_j 为交通小区 j 内包含的 POI 种类数.

居民社会经济属性和职住地建成环境属性的描述性统计结果如表 1 所示.

表 1 变量描述性统计结果
Table 1 Descriptive statistics of explanatory variables

变 量	变量描述	均 值	方 差
社会经 济属性	性 别	男为1,女为0	0.58
	年 龄	受访者年龄	36.60
	教育水平	大学及以上学历为1,其他为0	0.56
	家庭规模	家庭人口数量	3.00
	家庭年收入	<20 000 元为1,[20 000, 100 000)元为2,≥100 000 元为3	2.76
居住地建成 环境属性	到CBD距离	居住地到CBD欧几里得距离/km	5.23
	土地利用混合度	居住地交通小区内 11 类兴趣点数据混合度	0.60
	公共交通站点密度	居住地交通小区内公共交通站点密度/km ²	10.21
	交叉口密度	居住地交通小区内交叉口密度/km ²	26.92
工作地建成 环境属性	到CBD距离	工作地到CBD欧几里得距离/km	5.07
	土地利用混合度	工作地交通小区内 11 类兴趣点数据混合度	0.59
	公共交通站点密度	工作地交通小区内公共交通站点密度/km ²	10.78
	交叉口密度	工作地交通小区内交叉口密度/km ²	34.82

2 模型构建

因各影响因素度量方式(包括单位和数量级等)不同,传统的统计学模型难以估计各属性对因变量的作用程度.本文构建 GBDT 模型分析社会经济属性及职住地建成环境属性对小汽车拥有行为的影响程度.GBDT 模型是一种基于迭代的决策树算法,被广泛应用于捕捉各自变量对因变量的作用程度^[7-8].

在 GBDT 模型中,设 $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ 为包含 N 组自变量和因变量的训练集,其中, x_i 和 y_i 分别为第 i 组自变量和因变量.本文自变量包括:居民社会经济属性、居住地及工作地建成环境属性.因变量为二元变量,表示小汽车拥有行为,0 为没有小汽车,1 为拥有小汽车.GBDT 方法的核心为不断减小训练过程产生的残差,即

$$\hat{F}(x) = \arg \min_F \sum_{i=1}^N L[y_i, F(x)] \quad (2)$$

式中: $\hat{F}(x)$ 为待估计的近似函数; $L[y, F(x)]$ 为损失函数; y 为小汽车拥有行为变量; x 为社会经济属

性及职住地建成环境属性; $F(x)$ 为迭代计算得到的弱分类器估计值.

近似函数 $\hat{F}(x)$ 定义为弱分类器的迭代形式,即

$$\hat{F}(x) = \sum_{j=1}^J \gamma_j h_j(x) + c \quad (3)$$

式中: γ_j 、 c 为待估计参数; $h_j(x)$ 为弱分类器 j 的估计结果; J 为弱分类器数量.

通过梯度提升方法估计 GBDT 模型参数的计算过程如下:

Step 1 算法初始化,设置 $F_0(x)$ 为常数值, γ 为待估参数, $F_0(x) = \arg \min_{\gamma} \sum_{i=1}^N L(y_i, \gamma)$, 设置最大迭代次数 M .

Step 2 计算第 m 次迭代的残差, $r_{im} = -\left\{ \frac{\partial L[y_i, F(x_i)]}{\partial F(x_i)} \right\}_{F(x)=F_{m-1}(x)}$.

Step 3 利用弱分类器 $h_m(x)$ 拟合 Step 2 中得到的残差 r_{im} .

Step 4 估计乘子, $\gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L[y_i, F_{m-1}(x_i) + \gamma h_m(x_i)]$.

Step 5 更新, $F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$.

Step 6 判断是否达到迭代次数和精度要求. 满足, 则得到最终估计结果; 否则, 返回 Step 2.

GBDT模型通过计算所有加法树的均值, 量化捕捉外生变量对内生变量的影响程度, 即

$$\begin{cases} I_i^2 = \frac{1}{J} \sum_{j=1}^J I_i^2(T_j) \\ I_i^2(T_j) = \sum_{t=1}^{H-1} \hat{\tau}_i^2 I[v(t)=i] \end{cases} \quad (4)$$

式中: I_i^2 为对小汽车拥有行为的影响程度; T_j 为加法树 j ; $I_i^2(T_j)$ 为加法树 j 中自变量 x_i 对小汽车拥有行为的影响程度; t 为树的终端节点; H 为终端节点数量; $\hat{\tau}_i^2$ 为调整因子; $v(t)$ 为用于判断 i 是否属于该分割区域的函数; 当 i 属于该分割区域时, $I(\cdot)$ 值为1, 否则, $I(\cdot)$ 值为0.

为量化社会经济属性和职住地建成环境属性

对小汽车拥有行为的影响, 利用五折交叉验证方法确定模型参数. 数据被随机平均分为5份, 在执行过程中1份数据作为验证集合, 4份作为训练集合, 程序执行5次能够稳定估计结果; 将学习速率设置为0.001, 树的复杂度设置为10.

3 模型结果

各外生变量对小汽车影响程度的估计结果如表2所示. 由表2可知, 社会经济属性对小汽车拥有行为的影响程度最大, 为58.95%, 表明与职住地建成环境属性相比, 社会经济属性仍是影响小汽车拥有行为的最主要因素. 对于建成环境属性而言, 职住地建成环境对小汽车拥有行为的影响程度为41.05%, 其中, 居住地建成环境属性的影响程度为23.77%, 工作地建成环境的影响程度为17.28%. 尽管居住地建成环境属性对小汽车拥有行为的影响高于工作地, 在推进城市土地利用政策时, 对商业中心等就业岗位集中地区的土地利用规划亦是缓解小汽车增长的重要手段.

表2 各外生变量对小汽车拥有的影响程度

Table 2 Relative contribution of influential factors to explaining automobile ownership

变 量		影响程度排序	影响程度/%
社会经 济属性	性 别	10	3.23
	年 龄	3	10.67
	教育水平	2	13.98
	家庭规模	7	6.05
	家庭收入	1	25.02
居住地建成 环境属性	居住地到CBD距离	5	7.20
	居住地土地利用混合度	4	9.72
	居住地公共交通站点密度	8	5.83
	居住地交叉口密度	11	1.02
工作地建成 环境属性	工作地到CBD距离	9	4.04
	工作地土地利用混合度	6	6.43
	工作地公共交通站点密度	12	0.87
	工作地交叉口密度	7	5.94

社会经济属性中, 家庭收入、教育水平和年龄是对小汽车拥有行为影响程度较大的3个因素. 其中, 家庭收入是对小汽车拥有行为影响最大的变量, 这一结论与已有研究结果类似, 如Yin等^[4]针对中国48个城市的居民小汽车拥有行为进行建模发现, 在中国社会环境下家庭收入是影响家庭小汽车拥有行为的重要因素. 作为排名第2的变量, 教育水平对小汽车拥有行为的作用程度为13.98%, 也被

诸多研究所证实^[10]. 年龄对小汽车拥有行为影响程度排名第3, 即10.67%. 此外, 家庭规模和性别对小汽车拥有行为的作用程度分别为6.05%和3.23%.

居住地建成环境属性中, 除交叉口密度外, 其他建成环境因素对小汽车拥有行为的作用程度均高于对应的工作地建成环境属性. 其中, 土地利用混合度是所有居住地建成环境中对小汽车拥有行为影响最大的因素, 此结论为已有研究结论“土地

利用混合开发可减少小汽车依赖行为”提供了理论支持^[1].到CBD距离对小汽车拥有行为的影响程度仅次于土地利用混合度,为7.20%,表明居住地的空间位置是影响小汽车拥有行为的重要因素之一.公共交通站点密度对小汽车拥有行为的影响程度为5.83%.交叉口密度对小汽车拥有行为的作用较弱为1.02%.

工作地建成环境属性中,土地利用混合度仍是小汽车拥有行为最重要的影响因素.工作地交叉口密度对小汽车拥有行为的作用程度为5.94%,这也是工作地建成环境变量中唯一影响程度高于居住地建成环境的变量.此外,工作地到CBD距离及公共交通站点密度对小汽车拥有行为的作用程度分别为4.04%和0.87%.

由此可见,居民在小汽车拥有行为决策过程中,社会经济属性、居住地及工作地建成环境均发挥重要作用.

4 结 论

针对既有研究多忽略工作地建成环境对小汽车拥有的影响,本文基于GBDT模型对家庭小汽车拥有行为的影响因素进行建模,分析了职住地建成环境对小汽车拥有的作用程度.基于居民出行调查数据的实证研究表明,职住地建成环境均在小汽车拥有行为决策中发挥重要作用.本文可为通过城市规划手段减缓小汽车拥有提供理论基础.本文只考虑了各因素的作用程度,未能对各因素的影响方向进行分析,研究成果具有一定的局限性,未来在作用程度建模的基础上,分析各因素与小汽车拥有的非线性关系.

参考文献:

[1] 尹超英,邵春福,王晓全.考虑停车可用性的建成环境对小汽车通勤出行的影响[J]. 吉林大学报(工学版), 2018, 44(1): 1212-1221. [YIN C Y, SHAO C F, WANG X Q. Influence of urban built environment on car commuting considering parking availability[J]. Journal

of Jilin University (Engineering and Technology Edition), 2018, 44(1): 1212-1221.]

[2] 王晓全,邵春福,尹超英,等.考虑交通小区相关性的小汽车拥有行为研究[J]. 交通运输系统工程与信息, 2019, 19(4): 28-32, 71. [WANG X Q, SHAO C F, YIN C Y, et al. Capturing car ownership behavior considering spatial autocorrelation in traffic analysis zones[J]. Journal of Transportation Systems Engineering and Information Technology, 2019, 19(4): 28-32, 71.]

[3] DING C, WANG Y, YANG J, et al. Spatial heterogeneous impact of built environment on household auto ownership levels: Evidence from analysis at traffic analysis zone scales[J]. Transportation Letters, 2016, 8 (1): 26-34.

[4] YIN C, SUN B. Disentangling the effects of the built environment on car ownership: A multi-level analysis of Chinese cities[J]. Cities, 2018, 74: 188-195.

[5] GUO Z. Does residential parking supply affect household car ownership? The case of New York City[J]. Journal of Transport Geography, 2013, 26: 18-28.

[6] DING C, CAO X. How does the built environment at residential and work locations affect car ownership? An application of cross-classified multilevel model[J]. Journal of Transport Geography, 2019, 75: 37-45.

[7] DING C, CAO X, WANG Y. Synergistic effects of the built environment and commuting programs on commute mode choice[J]. Transportation Research Part A: Policy and Practice, 2018, 118: 104-118.

[8] 尹超英,邵春福,王聘玺,等.基于多层模型的城市建成环境对通勤行为的影响[J]. 交通运输系统工程与信息, 2018, 18(2): 122-127. [YIN C Y, SHAO C F, WANG P X, et al. Impacts of built environment on commuting behavior based on a multilevel modeling approach[J]. Journal of Transportation Systems Engineering and Information Technology, 2018, 18(2): 122-127.]

[9] MA X, DING C, LUAN S, et al. Prioritizing influential factors for freeway incident clearance time prediction using the gradient boosting decision trees method[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(9): 2303-2310.

[10] WANG X, SHAO C, YIN C, et al. Exploring the influence of built environment on car ownership and use with a spatial multilevel model: A case study of Changchun, China[J]. International Journal of Environmental Research and Public Health, 2018, 15: 1-15.