

引用格式:陈晓红,李洁月,杨艺,等.基于大规模用户数据的中国燃油私家车出行碳排放测算方法[J].资源科学,2024,46(4):717-727. [Chen X H, Li J Y, Yang Y, et al. Method for estimating carbon emissions from fuel-powered private cars in China based on large-scale user data[J]. Resources Science, 2024, 46(4): 717-727.] DOI: 10.18402/resci.2024.04.05

# 基于大规模用户数据的中国燃油私家车出行碳排放测算方法

陈晓红<sup>1,2,3</sup>, 李洁月<sup>1</sup>, 杨艺<sup>1,2</sup>, 胡东滨<sup>2,3</sup>

(1. 湖南工商大学前沿交叉学院, 长沙 410205; 2. 湘江实验室, 长沙 410205;  
3. 中南大学商学院, 长沙 410083)

**摘要:**【目的】私家车行驶碳排放是中国居民日常出行碳排放的主要来源,现有测算方法局限于宏观层面研究,本文以车主大数据和机器学习方法为基础,旨在构建燃油私家车出行碳排放量测算方法,可为道路交通碳排放量核算提供科学依据,助力实现碳达峰、碳中和。【方法】本文以2014—2023年中国13万余燃油车主披露的多维度数据为基础,运用线性回归、决策树、神经网络等方法,提出车辆(单车)累积出行碳排放量、相同行驶天数车辆群的平均累积出行碳排放量计算模型及预测模型,构建车辆年均出行碳排放量计算模型。【结果】①在车辆累积出行碳排放量预测方面,相比仅基于车龄的预测模型,由车型、制造商、区域、城市级别、价格区间等5个类别以及车龄构建的包含25个变量的模型可将预测精度( $R^2$ )由0.666提高到0.821,其中车龄对车辆累积出行碳排放量预测的影响程度仍然最高;②在相同行驶天数车辆群的平均累积出行碳排放量预测方面,基于神经网络方法的预测结果显示,全国层面的车辆平均累积出行碳排放量的预测精度达到0.915,车型、制造商、区域、城市级别、价格区间等5个类别的车辆平均累积出行碳排放量的预测精度处于0.875~0.925之间;③在车辆年均出行碳排放量测算方面,全国层面的车辆年均出行碳排放量为2.99 t,进口制造商、价格区间在(50, 100]万、价格高于100万等3个类别下的车辆年均出行碳排放量高于4 t,MPV、SUV车型、华南区域、超大城市级别、价格区间在(20, 30]万、价格区间在(30, 50]万等6个类别下的车辆年均出行碳排放量处于3.00~3.99 t之间,其余15个子类别划分下车辆年均出行碳排放量均低于全国层面的平均水平,处于2.15~2.99 t之间。【结论】具有车龄、车型、制造商、区域、城市级别、价格区间等多维特征的大规模车主数据以及神经网络模型有助于提高私家车使用周期碳排放量测算的科学性、全面性,不同区域的道路交通碳排放量测算应考虑上述特征差异。本文可为新能源私家车出行碳排放量测算研究提供参考。

**关键词:**燃油车;私家车;累积碳排放量;年均碳排放量;神经网络;预测

DOI: 10.18402/resci.2024.04.05

## 1 引言

全球气候变暖和碳排放已成为世界各国共同面临的严峻问题和关注焦点。2020年9月,习近平总书记在联合国大会提出“双碳”目标,进一步提高中国自主减排贡献力度。交通运输行业是碳排放的重要来源,是中国亟需控制碳减排的重点行业,

对碳达峰目标的实现有重要影响<sup>[1]</sup>。私家车碳排放约占交通运输部门碳排放量的50%,占据了最大份额<sup>[2]</sup>。中国低碳行动计划研究报告指出,2021年全国乘用车全生命周期碳排放量达7亿t,主要由汽油车产生,占比达到96%<sup>[3]</sup>。随着经济快速发展和人民生活水平不断提高,私家车数量不断上

收稿日期:2023-09-22;修订日期:2024-03-19

基金项目:国家自然科学基金基础科学中心项目(72088101);湖南省科技创新计划资助项目(2023RC3182)。

作者简介:陈晓红,女,湖南长沙人,中国工程院院士,教授,博士生导师,主要研究方向为资源能源环境管理、生态文明与绿色发展等。E-mail: c88877803@163.com

通讯作者:杨艺,男,湖南邵阳人,副教授,主要研究方向为大数据智能决策、低碳出行。E-mail: yangyihuc2018@163.com

升,燃油私家车碳排放已成为影响“双碳”目标和绿色低碳发展的重大挑战之一<sup>[4]</sup>,准确估计车辆出行碳排放量,可以为政府管理部门制定更加高效的碳排放调控政策提供决策参考,也是实现城市绿色交通和低碳发展的关键<sup>[5]</sup>。

私家车出行是居民日常出行的主要方式之一,其碳排放关联到家庭环境、交通环境、城市环境等多个方面,国内外学者们从不同视角探究了这些复杂环境对私家车出行碳排放的影响机理。基于乘用车碳排放和人口普查数据,Matsuhashi等发现人口密度大的地区由于居民拥有更多使用公共交通工具的机会,人均乘用车二氧化碳排放水平低于人口密度小的地区<sup>[6]</sup>。基于政府统计数据,Shiraki等发现车辆尺寸的增大和车辆保有量的增加导致了城市碳排放量的升高<sup>[7]</sup>。这些研究表明了人口分布、车辆保有量等城市特征以及车辆尺寸等车型特征对于碳排放都存在影响。Zhu等基于便携式排放测量系统对两款轻型燃油车的监测数据,识别出驾驶行为、发动机工况和外部环境是影响车辆碳排放的关键因素,并发现拥堵会造成更加激进的驾驶行为<sup>[8]</sup>。柴彦威等通过分析居民活动的调研数据,指出出行距离和方式是影响日常出行碳排放的主要因素,私家车的购置对居民家庭出行行为的高碳化具有不可逆的作用<sup>[9]</sup>。上述研究从城市特征、车型特征以及居民出行行为特征等不同角度开展的碳排放影响因素研究,为本文私家车出行碳排放量测算研究提供了基础。

近年来,尽管新能源汽车销量在不断增加,但燃油车保有量仍然占据市场主导地位。据公安部统计,2023年中国汽车保有量为3.36亿辆,新能源汽车占比6.07%,燃油车占比93.93%,私家车占比87.53%。燃油私家车的碳排放给环境带来很大的压力,对私家车碳排放量准确地测算成为了现实问题。目前,学者们主要以联合国政府间气候变化专门委员会(IPCC)“自下而上”法为基础,开展乘用车碳排放量的测算研究<sup>[10]</sup>。“自下而上”的碳排放量测算算法主要依托于对车辆数量、路网布局、交通流量、燃料消耗效率、行驶里程以及车辆行驶轨迹等关键因子的全面考量与精确分析,备受学者们的关注<sup>[11-16]</sup>。使用该方法,有学者分别以北京、深圳、济

南以及四大经济区域等中国城市或区域的私家车为研究对象,基于统计年鉴、居民问卷调查以及私家车轨迹等数据,构建了不同城市或区域的私家车碳排放量预测方法<sup>[1,11-14]</sup>。也有学者以日本、加拿大的私家车为研究对象,基于车企的年检数据、传感器采集的轨迹数据提出了私家车碳排放量测算模型<sup>[15,16]</sup>。值得注意的是,以私家车轨迹数据为驱动的研究中,神经网络、集成学习等智能技术逐渐被应用于测算碳排放量<sup>[12,16]</sup>。上述研究为本文燃油私家车碳排放量测算模型的构建提供了基础框架,但主要聚焦于部分城市或者经济区域,且依赖于大量详细且准确的数据采集,而当前的环境监测网络 and 手段存在显著不足,需要投入大量的人力、物力成本。

随着大数据和移动互联网的普及,Travelcard、汽车之家等国内外汽车在线服务平台为用户提供了便捷的汽车使用全生命周期的交流载体,这些平台产生的特征维度多、数据量规模大、公开化程度高的数据源得到了学者们的密切关注。有学者基于Spritmonitor.de、Travelcard、Geco air等国外平台提供的公开数据,开展了乘用车的官方油耗与车主真实油耗之间的差异研究<sup>[17]</sup>。有学者则以小熊油耗、汽车之家等国内平台的大规模公开数据为驱动,相继开展了中国各类乘用车的油耗差异以及乘用车行驶强度研究<sup>[18-21]</sup>。上述国内外研究验证了基于汽车在线服务平台的公开数据来开展油耗、行驶强度等碳排放相关研究的可行性,而乘用车的油耗、行驶强度(里程)是碳排放量测算的关键要素,这些研究为本文提供了较好的研究示范和基础。事实上,汽车在线服务平台不仅公布了油耗和里程数据,还深入披露了涵盖车主购车价格、购车地点、购车时间、车龄以及车型等多维度大规模数据,能够有效克服传统碳排放测算方法因数据采集所带来的人力与物力成本高昂的难题,有助于提高私家车碳排放量测算的科学性和有效性。

本文将以汽车在线服务平台披露的大规模数据为基础,以线性回归、决策树以及神经网络等机器学习模型为方法基础,研究车辆(单辆车)累积出行碳排放量计算模型、相同行驶天数车辆群的平均累积出行碳排放量计算模型、全国层面的年均出行

2024年4月

碳排放量计算模型以及按区域、城市级别、价格区间、车辆类别(尺寸)、制造商等不同类别划分下的车辆年均出行碳排放量计算模型,分析不同类别因素对乘用车出行碳排放量的影响,为全国乘用车碳排放总量估算、各类别划分下的乘用车碳排放总量估算提供方法支撑,为政府制定有针对性的交通减排方案提供参考依据。

## 2 数据来源和研究方法

### 2.1 数据来源

本文数据来自汽车之家服务平台2014年1月—2023年12月公开的16万余个燃油私家车车主数据(研究对象为燃油私家车,除特别说明,后文提到的车辆均为燃油私家车),涵盖中国31个省级行政区(港澳台地区因数据缺失未包含)344个城市。本文按照地理区域、车辆注册地的城市规模、销售价格范围、车辆类别(尺寸)和制造商对车辆进行分层抽样来减少样本偏差,使样本尽可能代表当前中国车辆市场的状况。按以上类别划分的乘用车占比,与2023年中国统计年鉴中的不同车型和地区的汽车市场份额相符,表明本文数据集符合市场分布。

(1)用户特征。在线服务平台汽车之家口碑模块的公开数据包括车主用户的ID、评论内容、行驶里程、百公里油耗、购买价格、购买时间、信息发布时间以及购买地点等信息。该平台还将用户划分为认证车主(上传车主行驶证及驾驶证)和非认证车主。

(2)车辆特征。汽车之家提供了丰富的车型库,包含了车辆型号、车辆制造商及车辆购买价格。车辆型号包括轿车、MPV和SUV3个类别,根据车辆制造商可以获得国产、进口和合资3个类别,车辆购买价格按照≤12万、(12, 20]万、(20, 30]万、(30, 50]万、(50, 100]万以及>100万分为6个级别。

(3)区域和城市特征。车主的区域归属主要根据31个省级行政区划分为7个区域(东北、华东、华中、华南、华北、西南和西北)。城市归属根据人口规模分为5类城市:小城市(包括Ⅰ型小城市和Ⅱ型小城市)、中等城市、大城市(包括Ⅰ型大城市和Ⅱ型大城市)、特大城市和超大城市<sup>[22]</sup>。

为了获取可信的有效样本数据,基于上述特征制定以下入选有效样本数据的规则:①认证车主;

②油耗范围(3.8~22.0 L/100 km);③累积行驶里程范围(20~500000 km);④日均行驶里程范围(1~519 km)。最终获取13万余个燃油私家车车主数据集。

### 2.2 研究方法

基于有效样本数据集以及神经网络等预测模型,构建出私家车(单辆车)日常出行累积碳排放量的计算模型及预测模型,挖掘累积碳排放的影响机理,提出全国层面以及车型、制造商、区域、城市级别、价格区间等级层面的平均累积碳排放量预测模型与年均碳排放量计算模型。从两方面开展研究:

(1)基于油耗和行驶里程构建每个车主样本的累积出行碳排放量计算模型,在大规模车主数据的驱动下,通过构建决策树、线性回归、神经网络等预测模型,分析车龄、车型、制造商、区域、城市级别、价格区间与车辆累积碳排放量之间的关系。

(2)在前述研究的基础上,提出一种相同行驶天数车辆群的平均累积出行碳排放量计算方法,进而构建出通用的全国层面以及5个类别层面的车辆年均碳排放量计算模型。

#### 2.2.1 车辆累积出行碳排放量计算模型

基于《中国汽车低碳行动计划》研究报告<sup>[23]</sup>中的碳排放测算方法,提出单辆车的累积出行碳排放量计算模型,具体定义如下。

定义1:车辆累积出行碳排放量计算模型如下:

$$C_{i,d} = FC_i \times K_{CO_2} \times T_{i,d} / 100 \quad (1)$$

式中: $C_{i,d}$ 为车龄(行驶天数)为 $d$ 的车辆样本 $i(i=1, 2, \dots, N_d)$ 的累积碳排放量(t); $N_d$ 为车龄为 $d$ (天)的样本总数; $FC_i$ 为样本 $i$ 的燃料消耗量(L/100 km); $K_{CO_2}$ 为转换系数(燃用汽油车型为2.37 kg/L); $T_{i,d}$ 为样本 $i$ 的行驶里程(km)。

为降低偏正态分布对车龄与累积碳排放量拟合精度的影响,本文采用对数处理方法,获取碳排放对数 $\ln C_{i,d}$ 和车龄对数 $\ln d$ ,并拟合线性回归模型 $\ln C_{i,d} = a \ln d + b$ ,结果如图1。未经过对数处理的碳排放量与车龄的拟合精度( $R^2$ )为0.666,基于对数处理的碳排放量与车龄的拟合精度提升至0.791。

上述分析结果显示,对数处理后的车龄与累积碳排放量存在较强的线性关系。此外,不同的车型、制造商、区域、城市级别和价格区间等对车辆的



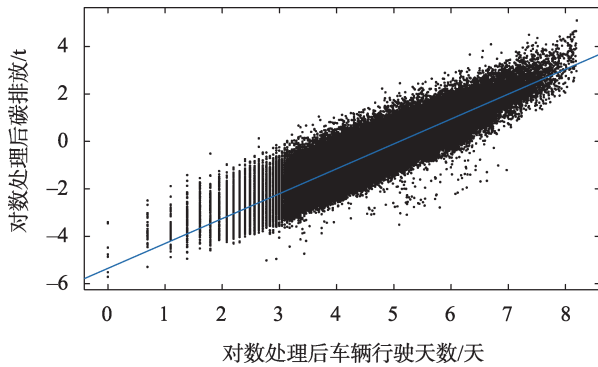


图1 基于对数处理的碳排放量与车龄拟合 ( $R^2=0.791$ )

Figure 1 Fitting carbon emissions to vehicle age based on logarithmic processing ( $R^2=0.791$ )

油耗、行驶率都存在不同程度的影响,进而影响车辆的累积碳排放量。因此,本文将考虑车型、制造商、区域、城市级别和价格区间等5个因素类别,并通过细分获取子类别(表1),结合车龄变量,构建出25个因变量的预测模型。

车辆累积碳排放量预测模型(M1):

$$\begin{cases} \ln C_{i,d} = f(\ln d_i, \psi_1^i, \psi_2^i, \psi_3^i, \psi_4^i, \psi_5^i) \\ i = 1, 2, \dots, 137928 \\ \psi_k^i = \{\psi_{kj}^i | j = 1, 2, \dots, J_k\} \\ \psi_{kj}^i = \begin{cases} 1, \text{样本} i \text{ 归属于因素} \psi_k \text{ 的子因素} \psi_{kj} \\ 0, \text{其他} \end{cases} \\ \sum_{j=1}^{J_k} \psi_{kj}^i = 1, \text{表示只属于因素} \psi_k \text{ 的1个子因素} \end{cases} \quad (2)$$

式中: $\psi_k^i$ 为样本*i*关于*k* ( $k = 1, 2, 3, 4, 5$ )个类别因素的子类别因素归属; $J_k$  ( $k = 1, 2, 3, 4, 5$ )为子类别个数,满足  $J = \{J_k | k = 1, 2, 3, 4, 5\} = \{3, 3, 7, 5, 6\}$ ,  $f$ 为预测模型。

表1 因素类别及其符号

Table 1 Factor categories and their symbols

因素类别	子类别	符号
车型	轿车、MPV、SUV	$\psi_1 = \{\psi_{1j}   j = 1, 2, 3\}$
制造商	国产、进口、合资	$\psi_2 = \{\psi_{2j}   j = 1, 2, 3\}$
区域	东北、华东、华中、华北、华南 西北、西南	$\psi_3 = \{\psi_{3j}   j = 1, 2, \dots, 7\}$
城市级别	小城市、中等城市、大城市、特大城市、超大城市	$\psi_4 = \{\psi_{4j}   j = 1, 2, \dots, 5\}$
价格区间	$\leq 12$ 万、(12, 20]万、(20, 30]万、 (30, 50]万、(50, 100]万、 $> 100$ 万	$\psi_5 = \{\psi_{5j}   j = 1, 2, \dots, 6\}$

通过如下例子来阐述每个样本对应的子类别归属情况。假设样本*i*为车主在东北的小城市购买的价格10万元的、国产汽车制造商生产的轿车,则该样本的24个子因素类别信息如下: $\psi_1^i = \{1, 0, 0\}$ ,  $\psi_2^i = \{1, 0, 0\}$ ,  $\psi_3^i = \{1, 0, 0, 0, 0, 0, 0\}$ ,  $\psi_4^i = \{1, 0, 0, 0, 0\}$ ,  $\psi_5^i = \{1, 0, 0, 0, 0, 0\}$ 。

基于定义1与车主有效样本集,获得模型M1中所有变量数据用于训练预测模型。采用线性回归、决策树以及神经网络等多种预测模型执行训练,具体结果如表2,其中神经网络模型对累积碳排放量的预测精度最高,达到0.821。

根据表2和图2可知,考虑子类别因素后,累积碳排放量的预测精度( $R^2$ )由0.791(单变量车龄)提升到0.821(25个变量),表明了各个子类别在一定程度上影响了累积碳排放量测算。为了获取每个子类别因素对累积碳排放量预测的影响程度,本文运用神经网络梯度法和多元回归特征重要度法获取子类别因素的影响程度,如图3所示。结果表明,车龄对累积碳排放量预测影响程度较高,即累积碳排放量在较大程度上依赖于车龄,价格、车型、区

表2 预测模型精度对比

Table 2 Comparison of the accuracy of multiple prediction models

模型 <i>f</i>	RMSE	$R^2$	时间/s	备注
线性回归	0.507	0.816	8	
树模型	0.555	0.780	82	精细树
神经网络	0.501	0.821	595	三层网络
SVM	0.502	0.820	3689	三次SVM

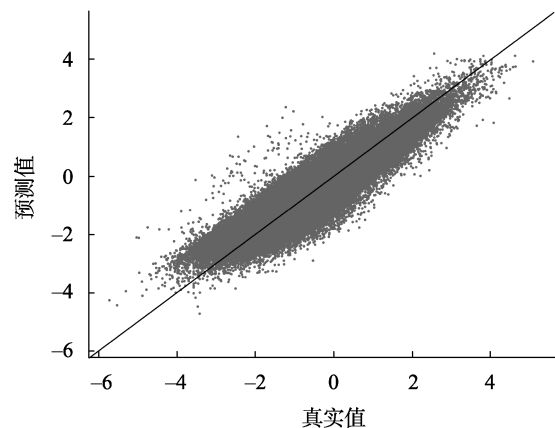


图2 累积碳排放量样本值与预测值 ( $R^2=0.821$ )

Figure 2 Sample and predicted values of cumulative carbon emissions ( $R^2=0.821$ )

2024年4月

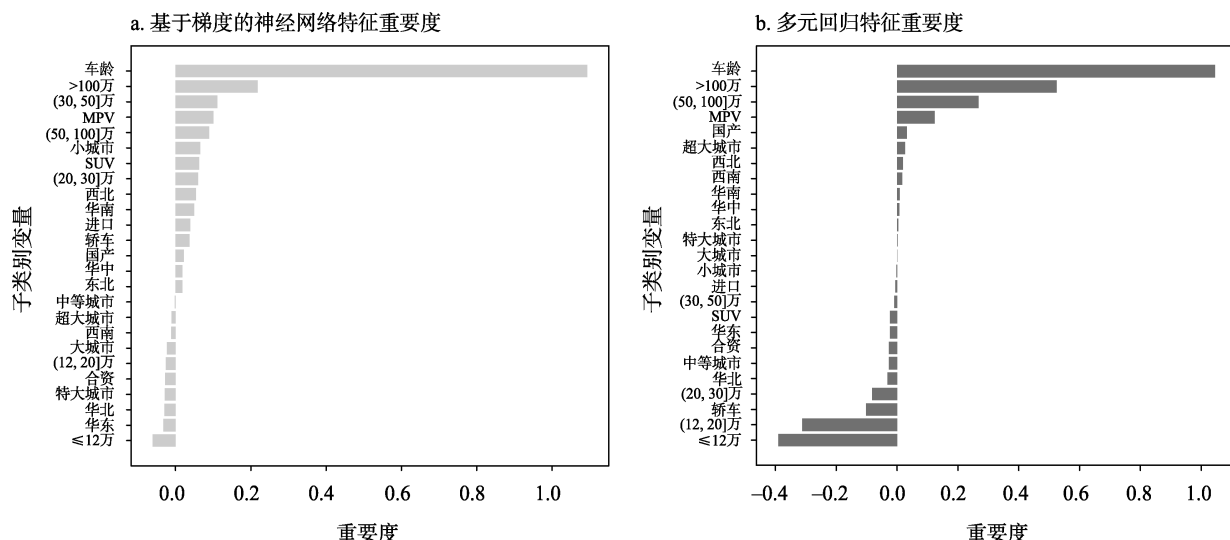


图3 子类别因素对累积碳排放量预测的影响程度:基于梯度的神经网络特征重要度(a)和多元回归特征重要度(b)

Figure 3 Degree of influence of subcategory factors on cumulative carbon emission prediction results: Gradient-based neural network feature significance (a) and multiple regression feature significance (b)

域、制造商以及城市级别等子类别因素都在一定程度上影响了累积碳排放量的预测,这为接下来的车辆年均出行碳排放量计算模型提供了科学依据。因此,在年均出行碳排放量计算模型的研究中,首先从全国层面构建车辆年均出行碳排放量模型,再从价格、车型、区域、制造商以及城市级别等不同层面探讨车辆碳排放量的计算模型。

### 2.2.2 车辆年均出行碳排放量计算模型

#### (1) 全国车辆年均出行碳排放量计算模型

不同于累积碳排放量的计算,由于同一行驶天数 $d$ 对应着多个车辆样本 $i(i=1,2,\dots,N_d)$ ,例如,总样本中有1000个车主样本的行驶天数为300天,则 $N_{300}=1000$ ,因此,需将行驶天数相同的群体样本的累积碳排放量聚合为综合的累积碳排放量,本文称之为平均累积出行碳排放量,以支撑车辆年均出行碳排放计算模型的构建。为了方便,记全体样本集合为 $G$ 。下面介绍平均累积出行碳排放量计算模型。

定义2:相同行驶天数车辆群的平均累积出行碳排放量计算模型:

$$C_d = e^{\sum_{i=1}^{N_d} \ln C_{i,d} / N_d} \quad (3)$$

式中: $C_d$ 为 $N_d$ 个群体样本聚合后的平均累积出行碳排放量。

行驶天数对应的样本分布呈现出偏正态分布,即不同行驶天数 $d$ 对应的样本量 $N_d$ 各有不同。基于

上述定义2,构建如下预测模型。

平均累积出行碳排放量预测模型(M2):

$$\ln C_d = f(\ln d) \quad (4)$$

式中: $f$ 为线性回归、决策树、神经网络等预测模型,通过预测精确度来确定模型。

基于上述预测模型获取平均累积出行碳排放量,构建下述定义中的全国车辆年均碳排放量计算模型。

定义3:全国车辆年均碳排放量 $\bar{C}$ 的计算模型为:

$$\bar{C} = w_1 C_{365} + \sum_{y=2}^{10} w_y (C_{365y} - C_{365(y-1)}) \quad (5)$$

式中: $\ln C_{365y} = f(\ln 365y)$ ;  $y=1,2,\dots,10$ ,其中 $y=1$ 表示2014年,依此类推, $y=10$ 表示2023年; $0 \leq w_y \leq 1$ 为各年份的市场份额权重,满足 $\sum_{y=1}^{10} w_y = 1$ ;依据预测模型M2来计算。

#### (2) 类别划分下车辆年均出行碳排放量计算模型

接下来,将构建车型、制造商、区域、城市级别以及价格区间等5个类别下的车辆年均出行碳排放量计算模型。为了方便,记 $G_j^k(j=1,2,\dots,J_k)$ 为归属于表1中第 $k$ 个类别 $\psi_k(k=1,2,3,4,5)$ 的子类别 $\psi_{kj}$ 样本集合,则有 $G = \bigcup_{j=1}^{J_k} G_j^k$ 。

在构建年均碳排放量计算模型之前,需要给出

每个子类别的平均累积碳排放量的计算模型。因此,在定义2的基础上,提出一个通用的子类别平均累积碳排放量计算模型。

定义4:子类别 $\psi_{kj}$ 群体样本集 $G_j^k$ 的平均累积出行碳排放量计算模型为:

$$C_d^{kj} = e^{\sum_{i=1}^{N_d^{kj}} \ln C_{i,d}^{kj} / N_d^{kj}} \quad (6)$$

式中: $C_{i,d}^{kj}$ 为车龄(行驶天数)为 $d$ 的车辆样本 $i$  ( $i=1, 2, \dots, N_d^{kj}$ )的累积出行碳排放量(t); $N_d^{kj}$ 为子类别 $\psi_{kj}$ 中车龄为 $d$ (天)的样本总数; $C_d^{kj}$ 为 $N_d^{kj}$ 个群体样本聚合后的平均累积出行碳排放量。

基于定义4,构建子类别 $\psi_{kj}$ 群体样本集 $G_j^k$  ( $j=1, 2, \dots, J_k$ )平均累积出行碳排放量预测模型。

子类别的平均累积出行碳排放量通用预测模型(M3):

$$\ln C_d^{kj} = f_{kj}(\ln d) \quad (7)$$

式中: $f_{kj}$ 为线性回归、决策树、神经网络等预测模型,通过预测精确度来确定模型。

定义5:子类别 $\psi_{kj}$ 的车辆年均碳排放量 $\bar{C}_{kj}$ 计算模型为:

$$\bar{C}_{kj} = w_1 C_{365}^{kj} + \sum_{y=2}^{10} w_y (C_{365y}^{kj} - C_{365(y-1)}^{kj}) \quad (8)$$

式中: $\ln C_{365y}^{kj} = f_{kj}(\ln 365y)$ ,  $y=1, 2, \dots, 10$ , 依据确定后的预测模型M3计算。

### 3 结果与分析

#### 3.1 全国车辆年均出行碳排放量计算

由定义3可知,全国车辆年均出行碳排放量计算模型是基于定义2中的平均累积碳排放量计算模型构建的。因此,首先分析全国层面的平均累积碳排放量情况。

根据定义2将全体13万余条样本聚合为2106条综合样本,即 $|C_d|=2106$ 。进一步地,运用决策树、线性回归、SVM以及神经网络等多种模型训练全国层面的平均累积碳排放量预测模型M2: $\ln C_d = f(\ln d)$ 。具体结果见表3和图4。由表3可知,相较于其他模型,神经网络模型的预测精度最高( $R^2=0.915$ )。图4展示了经过对数处理后,使用神经网络模型预测的平均累积碳排放量随车龄的变化情况。

根据2023年中国统计年鉴获取不同年份车龄的

表3 平均累积碳排放量不同预测模型 $f$ 精度对比

Table 3 Comparison of the accuracy of different prediction models for average cumulative carbon emissions

模型 $f$	RMSE	$R^2$	时间/s	备注
树模型	0.378	0.901	5	粗略树
线性回归	0.362	0.909	4	线性回归
SVM	0.360	0.910	3	粗略高斯SVM
神经网络	0.351	0.915	7	中型神经网络

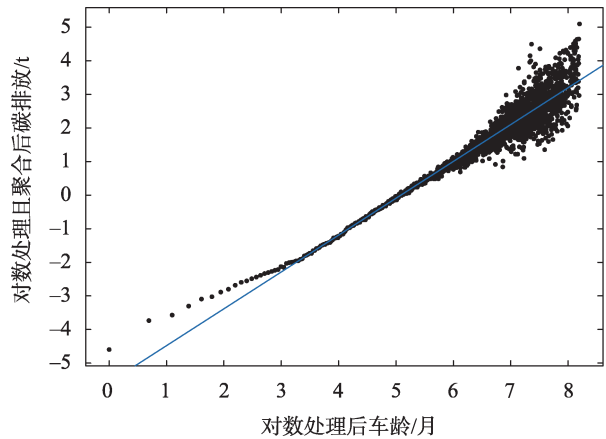


图4 基于神经网络得到的平均累积碳排放量与车龄的关系( $R^2=0.915$ )

Figure 4 Relationship between average cumulative carbon emissions and vehicle age based on logarithmic processing by neural network ( $R^2=0.915$ )

市场份额来计算累积碳排放量模型的权重向量 $W=(0.142, 0.141, 0.122, 0.113, 0.105, 0.096, 0.086, 0.076, 0.064, 0.055)^T$ ,其中元素代表2014—2023年的权重向量。

接下来,分析全国层面的车辆年均出行碳排放量的测算结果。运用拟合精度最高的预测模型获取各年份的累积碳排放量 $C_{365y}$  ( $y=1, 2, \dots, 10$ ),结合上述权重向量,根据定义3计算出全国车辆年均出行碳排放量 $\bar{C}=2.99$  t。

根据13万余车主有效样本集提供的行驶天数和定义1计算的累积碳排放量,可以换算出每个车主的年均碳排放量,全体样本车主年均碳排放量平均值为2.5 t,具体分布如图5所示。图5的X轴表示年均出行碳排放量,Y轴为累积分布概率。例如,点 $X=1.06$ 、 $Y=0.1$ 表示年均出行碳排放量 $\leq 1.06$  t的车主样本量占样本总体的10%。根据图5可得,80%的车主年均碳排放量处于(1.06, 4.30] t之间,60%的车主年均碳排放量处于(1.36, 3.37] t之间。由于行驶

2024年4月

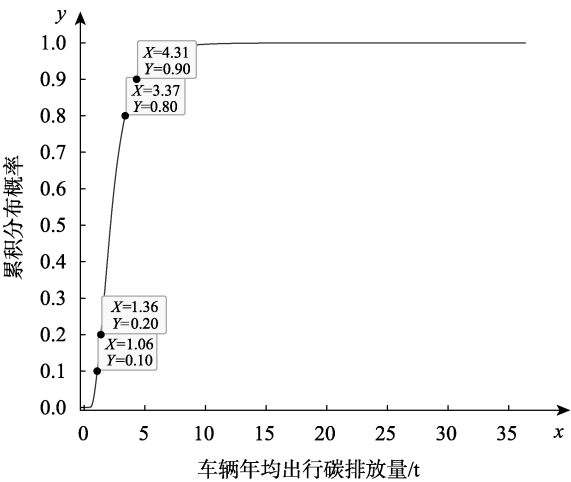


图5 按车龄换算的年均出行碳排放量累积概率

Figure 5 Cumulative probability of average annual carbon emissions from passenger car driving, converted by vehicle age

天数(车龄)相同的车主分布不同且呈现出偏正态分布,以基于车龄直接换算的年均碳排放量测算并

不合理。因此,本文采用在行驶天数相同的车主数据进行聚合(定义2,平均累积碳排放量)的基础上,进行年均碳排放量的测算,可以有效降低行驶天数的偏正态分布所带来的测算不精准的问题。

3.2 类别划分下车辆年均出行碳排放量计算

由定义4可知,类别划分下的年均出行碳排放量计算模型是基于定义2中平均累积碳排放量计算模型所构建。因此,首先分析类别划分下的相同行驶天数车辆群的平均累积出行碳排放量情况。

根据定义4和定义5,获取24个子类别划分下车辆平均累积碳排放量的拟合结果,预测模型主要采用决策树、线性回归、神经网络以及SVM等模型进行训练,选取精度最高的神经网络模型测算车辆平均累积碳排放量。子类别划分下平均累积碳排放量预测模型的预测精度和年均碳排放量测算结果见表4和图6。为了便于分析,表4给出了子类别

表4 基于神经网络模型的24个子类别的拟合结果

Table 4 Fitting results for 24 subcategories by neural network model

类别	子类别	样本量	样本占比/%	$R^2$	车辆年均碳排放量/t	平均行驶里程/km	平均油耗/(L/100 km)
车型	轿车	61537	44.61	0.914	2.56	5928.28	7.52
	MPV	10658	7.73	0.911	3.97	7106.17	9.36
	SUV	65735	47.66	0.922	3.24	5086.58	8.95
制造商	国产	39710	28.79	0.903	2.45	4944.23	8.34
	进口	6116	4.43	0.892	4.54	10033.49	10.25
	合资	92104	66.78	0.920	2.79	5615.52	8.22
区域	东北	7570	5.49	0.875	2.66	6009.24	8.51
	华东	50463	36.59	0.899	2.85	5387.49	8.37
	华中	19905	14.43	0.894	2.65	5355.96	8.27
	华北	19465	14.11	0.892	2.89	5820.29	8.33
	华南	15770	11.43	0.906	3.12	5982.18	8.38
	西北	8672	6.29	0.913	2.74	5663.42	8.07
城市级别	西南	16085	11.66	0.915	2.63	5856.34	8.40
	小城市	9171	6.65	0.898	2.56	5643.80	8.12
	中等城市	29101	21.10	0.907	2.57	5073.82	8.12
	大城市	48691	35.30	0.908	2.82	5501.24	8.29
	特大城市	26951	19.54	0.893	2.96	5787.92	8.46
	超大城市	24016	17.41	0.907	3.13	6314.50	8.68
价格/万	≤12	46510	33.72	0.920	2.15	5289.89	7.35
	(12, 20]	50134	36.35	0.925	2.54	5283.41	7.79
	(20, 30]	19350	14.03	0.920	3.44	6076.27	9.54
	(30, 50]	17595	12.76	0.914	3.70	5601.81	10.24
	(50, 100]	3541	2.57	0.900	4.87	8207.13	11.98
	>100	800	0.58	0.903	5.50	23500.24	13.96



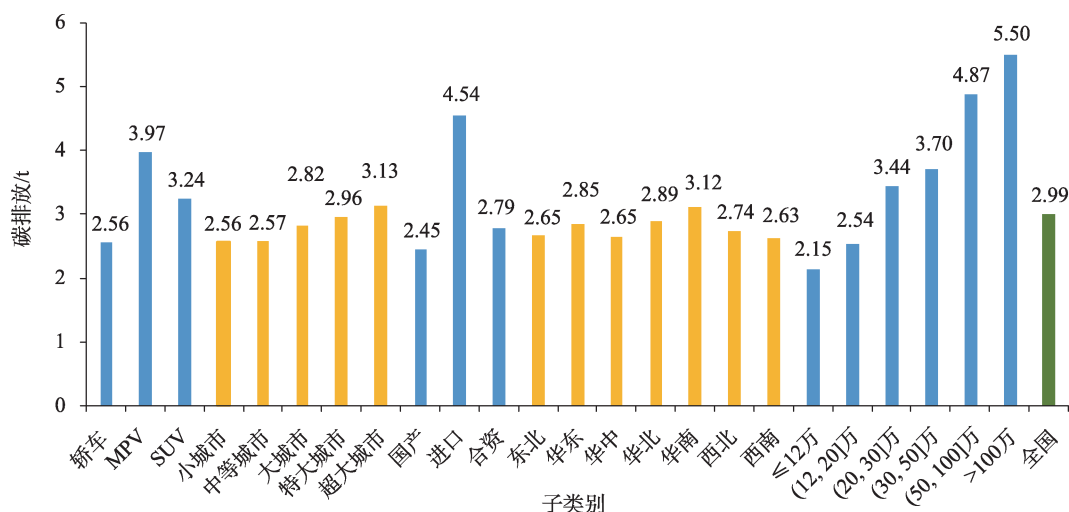


图6 全国层面与24个子类别的车辆年均碳排放量测算结果对比

Figure 6 Predicted average annual carbon emissions for China as a whole and by 24 subcategories

的样本量和占比(子类别样本量除以全体样本量)。

(1)车型类别划分下的车辆年均碳排放量。车辆年均碳排放量由高到低依次为MPV、SUV、轿车,MPV、SUV车型类别下的年均出行碳排放高于3 t,处于全国层面的平均水平之上,轿车车型则低于全国平均水平。根据定义1可知,碳排放随油耗递增。根据样本数据得出,这3类车型的油耗由高到低依次为MPV、SUV、轿车,与其碳排放量排序相同。

(2)制造商类别划分下的车辆年均碳排放量。进口制造商类别下的年均出行碳排放量高于4 t,合资和国产类别下的年均出行碳排放量低于全国层面的平均水平,分别为2.45 t和2.79 t。根据表4结果,相比国产车和合资车,进口车平均行驶里程、平均油耗均较高。

(3)区域类别划分下的车辆年均碳排放量。各个区域的车辆年均碳排放由高到低依次为华南、华北、华东、西北、东北、华中、西南。华南区域类别下的年均出行碳排放量高于3 t,其余区域类别划分下的年均碳排放量接近全国层面的平均水平,处于2.63~2.89 t之间。华南地区的道路交通较为发达,平均行驶里程较高,这可能是年均碳排放量高的主要原因。

(4)城市类别划分下的车辆年均碳排放量。城市级别越高,车辆年均碳排放量越大。超大城市类别下的年均出行碳排放量高于3 t,其余城市子类别划分下的年均碳排放量低于全国层面的平均水平,

处于2.56~2.96 t之间。其原因在于,不同级别的城市通勤时间和通勤距离存在差异。2023年度中国主要城市通勤监测报告显示,中国超大城市通勤时间达40 min,特大城市36 min,Ⅰ型大城市34 min;超大城市平均通勤距离为9.6 km,特大城市8.6 km,Ⅰ型大城市7.8 km。此外,城市通勤高峰交通拥堵程度可能也是影响不同级别城市车辆碳排放的原因,一般认为城市越大,交通拥堵越严重。

(5)价格类别划分下的车辆年均碳排放量。价格级别越高其对应的车辆年均碳排放量越高,(50, 100]万、高于100万类别下的年均出行碳排放量高于4 t,(20, 30]万、(30, 50]万类别下的年均出行碳排放量处于3.00~3.99 t之间,低于20万类别划分下的年均碳排放量低于全国层面的平均水平,处于2.15~2.54 t之间。一方面,由于车身重量及其高配置造成了油耗较高;另一方面,价位较高的平均行驶里程也相对较高。

## 4 结论与讨论

### 4.1 结论

本文以在线平台中大规模数据为驱动,以基于油耗、里程构建的累积碳排放量计算模型为基础,结合神经网络等机器学习方法构建了全国层面以及车型、制造商、区域、城市级别、价格区间等类别划分下的车辆平均累积出行碳排放量预测模型,测算了各层面下的车辆年均碳排放量。研究结论如下:



2024年4月

(1)基于25个子类别变量的车辆累积出行碳排放量预测模型结果表明,车型、制造商、区域、城市级别、价格区间等细粒度特征信息在一定程度上能够保证对车辆累积碳排放量的相对科学和有效的测算(预测精度0.821),其中车龄的影响程度最高,这也为基于车龄(行驶天数)构建的平均(按天)累积碳排放量预测模型提供了科学依据。

(2)以13万余车辆的累积出行碳排放量为基础,通过聚合行驶天数相同的车辆获取平均(按天)累积碳排放量,计算得到全国层面的车辆年均出行碳排放量为2.99 t,神经网络模型对于全国车辆平均累积出行碳排放量的预测精度达到0.915。

(3)车型、制造商、区域、城市级别、价格区间等类别划分层面,MPV和SUV车型、超大城市、进口制造商、华南区域、(20, 30]万、(30, 50]万、(50, 100]万以及>100万等9个类别下的车辆年均出行碳排放量高于3 t,其余15个子类别划分下的碳排放量处于2.15~2.99 t之间,低于全国层面的平均水平。另外,基于神经网络模型的平均(按天)累积碳排放量预测精度处于0.875~0.925之间,表明在车主大数据的驱动下,机器学习方法有助于提升车辆年均出行碳排放量的测算准确度。

(4)汽车在线服务平台公开披露的油耗、里程、购车价格、购车地点、购车时间、车龄以及车型等多维度特征的大规模数据,有效规避了传统测算方法中数据采集所带来的高昂人力与物力成本,提升了私家车碳排放量测算的科学性和有效性。在测算不同区域道路交通碳排放量时,深入剖析这些多维特征的差异,对于确保测算结果的精确性和实用性具有重要意义。

## 4.2 讨论

本文提出的燃油私家车日常出行累积碳排放量预测模型、年均碳排放量计算模型,一方面,可以为燃油私家车全生命周期碳排放量核算提供科学依据;另一方面,基于全国、车型、制造商、区域、城市级别、价格区间等各层面的燃油私家车保有量数据,可以为不同类别划分下所有燃油私家车碳排放总量的测算提供方法支撑。

随着人工智能技术的不断发展,相较于燃油车,新能源汽车的智能驾驶、车机交互、使用周期数

据监测等智能化水平更高。本文从大规模车主公开数据、神经网络等人工智能技术驱动视角下的碳排放量测算研究可以为新能源私家车碳排放量测算提供研究范式。

另外,本文仅分析了各类别下的车辆出行碳排放量,没有对各类别下的子类别执行多层次分类分析,还有待进一步的研究与探讨。在数据方面,本文并未获取每辆车的连续记录数据(速度、时间、定位等)数据集进行训练,后续将考虑结合在线服务平台车主数据和私家车轨迹数据开展研究,有助于提高累积碳排放量预测精度及年均碳排放量的测算准确度。

## 参考文献(References):

- [1] 王世进, 蒯乐伊. 中国交通运输业碳排放驱动因素与达峰路径[J]. 资源科学, 2022, 44(12): 2415-2427. [Wang S J, Kuai L Y. Driving factors and peaking path of CO<sub>2</sub> emissions for China's transportation sector[J]. Resources Science, 2022, 44(12): 2415-2427.]
- [2] International Energy Agency. Key World Energy Statistics[R/OL]. (2019-12-01)[2023-09-10]. <https://www.iea.org/topics/transport>.
- [3] 中汽数据有限公司. 中国汽车低碳行动计划(2022)[R]. 天津: 中汽数据有限公司, 2022. [Automotive Data of China. China Automobile Low Carbon Action Plan Research Report (2022)[R]. Tianjin: Automotive Data of China, 2022.]
- [4] Zhang W, Lawell C L, Umanskaya V I. The effects of license plate-based driving restrictions on air quality: Theory and empirical evidence[J]. Journal of Environmental Economics & Management, 2017, 82: 181-220.
- [5] Milovanoff A, Posen I D, Maclean H L. Electrification of light-duty vehicle fleet alone will not meet mitigation targets[J]. Nature Climate Change, 2020, 10(12): 1-6.
- [6] Matsuhashi K, Ariga T. Estimation of passenger car CO<sub>2</sub> emissions with urban population density scenarios for low carbon transportation in Japan[J]. IATSS Research, 2016, 39(2): 117-120.
- [7] Shiraki H, Matsumoto K, Shigetomi Y, et al. Factors affecting CO<sub>2</sub> emissions from private automobiles in Japan: The impact of vehicle occupancy[J]. Applied Energy, 2020, DOI: 10.1016/j.apenergy.2019.114196.
- [8] Zhu X H, Lu K F, Peng Z R, et al. Characterizing carbon emissions from China V and China VI gasoline vehicles based on portable emission measurement systems[J]. Journal of Cleaner Production, 2022, DOI: 10.1016/j.jclepro.2022.134458.
- [9] 柴彦威, 肖作鹏, 刘志林. 居民家庭日常出行碳排放的发生机制与调控策略: 以北京市为例[J]. 地理研究, 2012, 31(2): 334-344. [Chai Y W, Xiao Z P, Liu Z L. Low-carbon optimization strat-

- egies based on CO<sub>2</sub> emission mechanism of household daily travels: A case study of Beijing[J]. *Geographical Research*, 2012, 31(2): 334–344.]
- [10] Intergovernmental Panel on Climate Change. 2006 IPCC Guidelines for National Greenhouse Gas Inventories[M]. Japan: Institute for Global Environmental Strategies (IGES), 2006.
- [11] Li P H, Jones S. Vehicle restrictions and CO<sub>2</sub> emissions in Beijing: A simple projection using available data[J]. *Transportation Research Part D: Transport and Environment*, 2015, 41: 467–476.
- [12] 陈文婕, 吴小刚, 肖竹. 中国四大经济区域道路交通碳排放预测与减排潜力评估: 基于私家车轨迹数据的情景模拟[J]. *经济地理*, 2022, 42(7): 44–52. [Chen W J, Wu X G, Xiao Z. Carbon emission prediction and emission reduction potential assessment of road traffic in China's four major economic regions: Scenario simulation based on private vehicle trajectory data[J]. *Economic Geography*, 2022, 42(7): 44–52.]
- [13] Zeng Y, Tan X, Gu B, et al. Greenhouse gas emissions of motor vehicles in Chinese cities and the implication for China's mitigation targets[J]. *Applied Energy*, 2016, 184: 1016–1025.
- [14] Tan X C, Tu T Q, Gu B H, et al. Scenario simulation of CO<sub>2</sub> emissions from light-duty passenger vehicles under land use-transport planning: A case of Shenzhen International Low Carbon City[J]. *Sustainable Cities and Society*, 2021, DOI: 10.1016/j.scs.2021.103266.
- [15] Matsuhashi K, Ariga T, Ishikawa M. Estimation of passenger car CO<sub>2</sub> emissions by population density class based on Japanese vehicle inspection certificate data[J]. *IATSS Research*, 2023, 47(2): 179–184.
- [16] Natarajan Y, Wadhwa G, Sri Preethaa R, et al. Forecasting carbon dioxide emissions of light-duty vehicles with different machine learning algorithms[J]. *Electronics*, 2023, DOI: 10.3390/electronics12102288.
- [17] Ktistakis M A, Pavlovic J, Fontaras G. Developing an optimal sampling design to monitor the vehicle fuel consumption gap[J]. *Science of the Total Environment*, 2022, DOI: 10.1016/j.scitotenv.2022.154943.
- [18] Dror M B, Qin L Z, An F. The gap between certified and real-world passenger vehicle fuel consumption in China measured using a mobile phone application data[J]. *Energy Policy*, 2019, 128: 8–16.
- [19] Li Y W, Zeng I Y, Niu Z H, et al. Predicting vehicle fuel consumption based on multi-view deep neural network[J]. *Neurocomputing*, 2022, 502: 140–147.
- [20] Yu R J, Ren H H, Liu Y, et al. Gap between on-road and official fuel efficiency of passenger vehicles in China[J]. *Energy Policy*, 2021, DOI: 10.1016/j.enpol.2021.112236.
- [21] Ou S Q, Yu R J, Lin Z H, et al. Intensity and daily pattern of passenger vehicle use by region and class in China: Estimation and implications for energy use and electrification[J]. *Mitigation and Adaptation Strategies for Global Change*, 2020, 25: 307–327.
- [22] 中华人民共和国国家发展和改革委员会. 关于调整城市规模划分标准的通知[EB/OL]. (2014–10–29) [2023–09–10]. [https://www.ndrc.gov.cn/xwdt/ztl/xxczhjs/ghzc/201605/t20160509\\_971910.html](https://www.ndrc.gov.cn/xwdt/ztl/xxczhjs/ghzc/201605/t20160509_971910.html). [National Development and Reform Commission. Circular on the Adjustment of the Criteria for the Classification of the Size of Cities[EB/OL]. (2014–10–29) [2023–09–10]. [https://www.ndrc.gov.cn/xwdt/ztl/xxczhjs/ghzc/201605/t20160509\\_971910.html](https://www.ndrc.gov.cn/xwdt/ztl/xxczhjs/ghzc/201605/t20160509_971910.html).]
- [23] 中汽数据有限公司. 中国汽车低碳行动计划(2021)[R]. 天津: 中汽数据有限公司, 2021. [Automotive Data of China. China Automobile Low Carbon Action Plan Research Report (2021)[R]. Tianjin: Automotive Data of China, 2021.]

## Method for estimating carbon emissions from fuel-powered private cars in China based on large-scale user data

CHEN Xiaohong<sup>1,2,3</sup>, LI Jieyue<sup>1</sup>, YANG Yi<sup>1,2</sup>, HU Dongbin<sup>2,3</sup>

(1. School of Advanced Interdisciplinary Studies, Hunan University of Technology and Business, Changsha 410205, China;

2. Xiangjiang Laboratory, Changsha 410205, China; 3. School of Business, Central South University, Changsha 410083, China)

**Abstract: [Objective]** The objective of this study was to develop a precise calculation method for fuel-powered private car travel carbon emissions, utilizing big data from car owners and employing machine learning techniques. This method aims to address the limitations of existing macro-level

studies and provide a scientific basis for accurately estimating road transport carbon emissions. It will also contribute to achieving peak carbon dioxide emissions and promoting carbon neutrality in China. **[Methods]** Based on multidimensional data disclosed by over 130000 fuel vehicle owners in China from 2014 to 2023, this article utilizes methods such as linear regression, decision trees, and neural networks to propose a calculation model and prediction model for the average cumulative travel carbon emissions of vehicle groups with the same number of days traveled as well as individual vehicles. Subsequently, a calculation model for annual average travel carbon emissions of vehicles is constructed. **[Results]** (1) In terms of predicting the cumulative travel carbon emissions of vehicles, compared to the vehicle cumulative travel carbon emission prediction model based solely on vehicle age, the prediction accuracy ( $R^2$ ) of the model with 25 variables in five categories (vehicle type, manufacturer, region, city level, price range) and vehicle age increased from 0.666 to 0.821. Age was found to have the highest impact on predicting cumulative travel carbon emissions for vehicles. (2) In terms of predicting the average cumulative carbon emissions from vehicles in a group with the same number of days traveled, the results based on neural network methods show that the prediction accuracy for the average cumulative travel carbon emissions at a national level reaches 0.915. The prediction accuracy for average cumulative travel carbon emissions for vehicles in five categories including vehicle type, manufacturer, region, city level, and price range ranges from 0.875 to 0.925. (3) In terms of calculating the average annual carbon emissions from vehicle travel, the national level has an average of 2.99 t. Vehicles in three categories—imported manufacturers, prices between (500, 1000] thousand CNY, and prices over 1000 thousand CNY—have higher average annual carbon emissions exceeding 4 t. Vehicles in six categories—MPV and SUV models, South China region, super large cities, price range between (200, 300] thousand CNY and price range between (300, 500] thousand CNY—have average annual carbon emissions ranging from 3.00 to 3.99 t. The remaining fifteen subcategories have lower average annual carbon emissions than the national level's average level and fall within the range of 2.15 to 2.99 t. **[Conclusion]** The large-scale vehicle owner data of multiple dimensions such as vehicle age, vehicle type, manufacturer, region, city level, and price range, along with the neural network model, contribute to improving the scientific and comprehensive measurement of carbon emissions from private car usage cycles. The calculation of road traffic carbon emissions in different regions should take into account the aforementioned differences in features. This study provides a reference for calculating carbon emissions from new energy private car travel.

**Key words:** fuel-powered cars; private cars; cumulative carbon emissions; average annual carbon emissions; neural networks; prediction