

# 基于MI特征选择的车辆能耗高精度预测方法

王 宁<sup>1</sup>, 李秀峰<sup>1</sup>, 聂辽栋<sup>1</sup>, 刘登程<sup>2</sup>, 于 勤<sup>3</sup>, 樊华春<sup>3</sup>, 徐 炜<sup>3</sup>

(1. 同济大学汽车学院, 上海 201800; 2. 南昌智能新能源汽车研究院, 南昌 330052; 3. 江西五十铃汽车有限公司, 南昌 330199)

**摘要:** 近年来,机器学习方法在车辆实时能耗预测方面得到了广泛应用,但实车采集数据中存在的精度不足、字段缺失以及多重共线性等问题,尤其是同款车型中驾驶工况和驾驶者行为存在显著差异,限制了能耗预测准确性和泛化能力的进一步提升。为此,本文系统考虑特征冗余度、数据平衡性、货运趟次、运输能力、路段拥挤程度和司机驾驶时长等因素,使用交互信息(MI)方法选择关键特征,并构建司机特征画像作为独立特征,进而结合XGBoost、RF和MLP等机器学习方法提出一种基于MI特征选择的能耗高精度预测方法,然后基于120辆轻型卡车的T-BOX采集数据进行实例验证。结果表明,本文提出的预测方法能够显著提高不同驾驶行为和驾驶工况下的能耗预测精度,研究成果可为开发预测轻卡能耗的通用模型提供参考。

**关键词:** 车辆能耗预测; 轻型卡车; 交互信息方法; 司机特征画像; 机器学习

中图分类号: U461.8

文献标志码: A

## High-Precision Vehicle Energy Consumption Prediction using Mutual Information Feature Selection

WANG Ning<sup>1</sup>, LI Xiufeng<sup>1</sup>, NIE Liaodong<sup>1</sup>, LIU Dengcheng<sup>2</sup>, YU Qin<sup>3</sup>, FAN Huachun<sup>3</sup>, XU Wei<sup>3</sup>

(1. School of Automotive Studies, Tongji University, Shanghai 201800, China; 2. Nanchang Automotive Institute of Intelligence & New Energy, Nanchang 330052, China; 3. Jiangxi Isuzu Motors Co., Ltd., Nanchang 330199, China)

**Abstract:** In recent years, machine learning methods have been widely adopted for real-time vehicle energy consumption predictions. However, the accuracy and generalizability of these predictions are often hindered by challenges such as data imprecision, missing fields, multicollinearity, and substantial difference in driving conditions and driver behaviors among identical vehicle models. To address these issues, this study systematically

considers factors such as feature redundancy, data balance, freight trip frequency, transport capacity, traffic congestion and driving duration. Subsequently, an energy consumption prediction model with high precision is developed using a combination of machine learning methods such as XGBoost, Random Forest (RF), and Multilayer Perceptron (MLP). The model utilizes key features selected through the Mutual Information (MI) method, along with a constructed driver profile that captures characteristic behaviors as an independent feature. The proposed method is validated using T-BOX data collected from 120 light trucks. Experimental results indicate that the prediction method significantly enhances the prediction accuracy of energy consumption under various driving behaviors and conditions. This research contributes to the development of models with high precision in estimating the fuel consumption of light trucks.

**Keywords:** vehicle energy consumption prediction; light trucks; mutual information method; driver profile; machine learning

目前,世界各国大多数机动车仍然依赖于石油等不可再生能源,其在运行过程中会排放大量的有害物质。面对石油资源逐渐短缺和环境质量不断恶化的境况,不仅需要加快新能源车辆的研发和推广,如何降低燃油车的燃料消耗也是当务之急。在这一背景下,轻卡逐渐成为了人们的关注焦点。这种车型主要用于500 km以内的短途运输,也是国内销量和保有量最高的货车种类。然而,公路货运卡车燃油消耗与其本身特性、行驶工况、货运种类、装载量等因素密切相关,因此精准的能耗预测对于提高运输效率、降低货运成本以及节能减排具有重要的意义。特别是在交通拥堵的情况下,油耗和排放量会

收稿日期: 2023-10-23

基金项目: 南昌智能新能源汽车研究院科研项目(TPD-TC202303-11)

第一作者: 王宁(1977—),男,副教授,博士生导师,管理学博士,主要研究方向为汽车大数据分析。E-mail: wangning@tongji.edu.cn

通信作者: 聂辽栋(1998—),男,博士生,主要研究方向为汽车大数据分析。E-mail: 2311436@tongji.edu.cn

因为平均速度降低和加减速频繁而增加。同时,轻卡行业目前面临大吨小标治理及合规化等要求<sup>[2]</sup>,预计单车运力将下降 35%~50%,这预示着随着销量的持续增长,轻卡能耗的研究变得尤为关键。

近年来,不断有学者致力于车辆油耗模型的研究,早期的研究主要基于车辆动力学原理,其中以车辆比功率(VSP)模型、VT-Micro模型和综合模态排放模型(CMEM)为代表<sup>[3-5]</sup>,但这类模型需要深入了解车辆内部构造与发动机原理,考虑因素太多,并且模型校准过程复杂,当输入规模变大时,运算效率将会随之降低。随着数据挖掘和信息技术的不断进步,机器学习方法在油耗预测中已经得到了广泛应用<sup>[6-7]</sup>。相较于传统的物理模型,机器学习模型的优势在于能够高效地处理大规模样本数据。特别是利用车载 T-Box 设备收集的实时数据,这允许对模型进行即时的算法更新和迅速迭代。

赵晓华<sup>[8]</sup>通过提取出租车司机的微观驾驶行为数据,利用主成分分析和 BP 神经网络,成功建立了城市快速路基础路段的油耗预测模型。而 Zhigang Xu & Tao Wei<sup>[9]</sup>利用车联网数据,提出了基于能源消耗指数(ECI)的半物理模型与广义回归神经网络(GRNN)的数据驱动模型,并同时与 VSP、VT-Micro 和 CMEM 模型进行比对,结果显示所提出的模型在预测卡车燃料消耗上展现出更强的性能。Sandareka Wickramanayake<sup>[10]</sup>采用 GPS 跟踪装置和电容式高精度燃油传感器数据,通过随机森林(RF)、梯度提升(GB)和神经网络方法建立长途公共巴士的油耗预测模型。Ying Yao<sup>[11]</sup>则从另一个角度出发,利用智能手机收集的出租车司机驾驶行为数据,与车载诊断系统(OBD)的油耗数据相结合,利用 BP 神经网络、支持向量回归机和随机森林构建出租车驾驶员个体油耗预测模型。另外,Yiman Du<sup>[12]</sup>采用 BP 神经网络,从时间和空间两个维度全面研究油耗与各种影响因素之间的关系,研究表明,这种模型在预测燃料消耗方面具有较好的性能。黄赫<sup>[13]</sup>利用 SelectKbest 函数对 OBD 采集的数据进行分析,并结合 Tensorflow 的多层感知机(MLP)神经网络模型及支持向量机(SVM)进行油耗预测。但需要注意的是,该数据是在冬季低温工况下采集的,这可能会对预测结果产生一定的偏差。而邹智宏<sup>[14]</sup>则结合最大相关最小冗余算法与主成分分析,去除数据的冗余性,并利用 BP 神经网络为高速公路车辆油耗进行建模。不过,在高速公路上进行燃料消耗建模相对较为简单,因为如交通和道路状况等外部因素在大

部分时间内都相对稳定。

为提高车辆能耗预测准确性和模型泛化能力,本文针对轻型柴油卡车驾驶工况和驾驶者行为差异,构建司机特征画像作为独立特征。同时,本文使用交互信息(mutual information, MI)方法确定车辆运行对油耗影响的关键特征,再考虑行驶路段的拥挤程度,以此建立一种基于 MI 特征选择的能耗预测模型。最后,通过 120 辆轻卡的 T-BOX 采集数据进行模型验证。结果表明,本研究提出的能耗预测方法能够显著提高不同驾驶行为和驾驶工况下的预测精度,研究成果可以为开发预测轻卡能耗的通用模型提供参考。

## 1 数据来源与特征筛选

### 1.1 实时运行状态数据采集

本研究所用数据样本是通过车载数据采集终端(T-BOX)固件中的采集功能获取的轻型柴油卡车运行的数据集,在中国多个省市收集了 120 辆技术参数相同的轻型柴油卡车的行驶数据,提取驾驶员驾驶行为及车辆运行数据,车辆数据的采样间隔为 10 s,数据格式如表 1 所示。

由于实际数据信息的采集与传输过程中,部分数据会出现异常或缺失现象,需要对原始数据进行数据清洗。主要包括时间跳变检查、缺失数据填充、对不符合实际情况的数据进行调整或剔除等,提高数据质量为后续建模等进行准备。

### 1.2 MI 算法与特征筛选

交互信息是一种度量两个随机变量之间的相互依赖性的方法,与相关系数不同,交互信息不仅可以捕获线性关系,还可以捕获更复杂的非线性关系。在机器学习和数据分析中,特征选择是提高模型性能和减少训练时间的关键步骤。特别是在面对高维数据时,选择与目标变量高度相关的特征可以大大提高模型的准确性。MI 可以通过以下公式定义:

$$MI(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

式中:  $MI(X, Y)$  是随机变量  $X$  和  $Y$  之间的交互信息;  $p(x, y)$  是  $X$  和  $Y$  的联合概率分布;而  $p(x)$  和  $p(y)$  分别是  $X$  和  $Y$  的边缘概率分布。

交互信息衡量的是通过观察  $X$  而获得  $Y$  的信息量。如果两个变量完全独立,则它们之间的交互信息为零。如果  $X$  和  $Y$  之间存在某种关系,则交互

表 1 TBOX 数据格式  
Tab.1 Data format of TBOX

车辆识别码	采集时间	经度	纬度	车速	大气压力	发动机净输出扭矩	摩擦扭矩	发动机转速	发动机燃料流量
LETXXX50	1/12/2022 00:04:04	121.1392°	31.28805°	66.648 km/h	103 kPa	55 %	6 %	1638.5 r/min	10.55 L/h
LETXXX50	1/12/2022 00:04:14	121.1390°	31.28956°	68.578 km/h	103 kPa	48 %	6 %	1691.0 r/min	9.30 L/h
...	...	...	...	...	...	...	...	...	...
车辆识别码	反应剂余量	进气量	SCR 入口	SCR 出口温度	发动机冷却液温	累计里程	发动机状	续航里程	剩余油量
LETXXX50	41.6 %	184.3 kg/h	296.219 °C	286.3125 °C	77 °C	26 463 km	开启	436 km	65.2 %
LETXXX50	41.6 %	174.7 kg/h	292.125 °C	288.8125 °C	79 °C	26 464 km	开启	435 km	65.2 %
...	...	...	...	...	...	...	...	...	...

信息将大于零,且关系越强,交互信息的值越大。

对燃油消耗影响因素研究中,显示影响车辆能耗排放的因素众多,国内外文献<sup>[10-11]</sup>主要从汽车发动机参数与汽车行驶参数出发选取 10~16 个特征数量作为影响因素。本研究对油耗预测输入特征进行平均值、标准差、最大值及最小值计算,再通过 MI 方法量化每个特征与油耗变量之间的依赖性,对输入数据进行筛选,最后选择 MI 评分前 10 名作为油耗预测输入特征。由式(1)对子集特征的重要性进行计算,排序结果如表 2 所示。

表 2 油耗预测输入特征的交互信息评分与排序  
Tab.2 MI Scores and ranking of fuel consumption prediction input features

油耗预测输入特征	MI 评分	排序
发动机燃料流量平均值	1.602 504	1
进气量平均值	1.557 927	2
车速平均值	1.486 842	3
发动机净输出扭矩平均值	1.463 839	4
发动机转速标准差	1.425 313	5
发动机燃料流量标准差	1.418 684	6
发动机转速平均值	1.413 199	7
车速标准差	1.402 512	8
车速最小值	1.378 864	9
发动机净输出扭矩标准差	1.372 777	10

## 2 司机驾驶行为分类

### 2.1 驾驶行为特征参数构建

司机驾驶行为是影响车辆能耗的关键因素之一。在传统的驾驶行为研究中,分析重点往往放在与车辆实时工况如车速、发动机扭矩等紧密相连的特征上。然而,这种方法往往忽略了一些与驾驶行为直接相关的非工况统计量。为了更全面地理解和刻画这种行为模式,本研究引入了一系列描述司机驾驶习惯和运输模式的统计量,如表 3、表 4 所示。这些统计量涵盖了从驾驶时段到总行驶里程等多个维度,提供了对司机驾驶行为的宏观和微观视角。

表 3 司机驾驶行为特征变量  
Tab.3 Driver transportation behavior feature variables and abbreviations

特征变量	单位	符号	特征变量	单位	符号
运营天数	—	$D$	总行驶里程	km	$M_T$
总运营时长	h	$T_O$	日均行驶里程	km	$M_{TD}$
日均运营时长	h	$T_{OD}$	夜间驾驶总时长	h	$N_{TD}$
总停留时长	h	$T_S$	夜间驾驶时长比例	—	$N_{TR}$
日均停留时长	h	$T_{SD}$	夜间驾驶频率	—	$N_{DF}$
总运营效率	—	$E_O$	疲劳驾驶频率	—	$F_{DF}$

表 4 司机运输行为特征变量  
Tab.4 Driver transportation behavior feature variables

车辆识别码	$D$	$T_O$	$T_{OD}$	$T_S$	$T_{SD}$	$E_O$	$M_{start}$	$M_{end}$	$M_T$	$M_{TD}$	$N_{TD}$	$N_{TR}$	$N_{DF}$	$F_{DF}$
LETXXX43	31	100.275	3.234 677	38.191 67	1.231 989	0.724 182	55 124	57 975	2 851	91.967 74	0	0	0	0
LETXXX12	30	394.019	13.133 98	19.994 44	0.666 481	0.951 706	17 933	21 723	3 790	126.333 3	118.844 4	0.301 621	0.966 667	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

统计数据如"总运营时长"、"运营天数"和"总运营效率"等,能够展示司机的工作节奏和驾驶时长。而与风险相关的统计指标,如"夜间驾驶总时长"、"夜间驾驶频率"和"疲劳驾驶频率",则表征了司机的驾驶风险驾驶倾向。Grandjean 发现,疲劳是与驾驶性能下降相关的渐进过程,需要通过适当休息来缓解<sup>[15]</sup>。Kaneko & Jovanis (1992)指出,夜间驾驶与事故风险显著相关,尤其驾驶时长超过 9 h 以上时<sup>[16]</sup>。

$$D = D_{end} - D_{start} \tag{2}$$

式中:  $D_{start}$ 、 $D_{end}$  分别为车辆运行开始和结束日期。

$$T_O = \sum_{i=1}^m t_i \tag{3}$$

式中:  $t_i$  为第  $i$  次运行的时间长度;  $m$  为运行的总次数。

$$T_{OD} = \frac{T_O}{D} \tag{4}$$



$$T_s = \sum_{i=1}^m s_i \quad (5)$$

式中:  $s_i$  为第  $i$  次停留的时间长度。

$$T_{SD} = \frac{T_s}{D} \quad (6)$$

$$E_o = \frac{T_o}{T_o + T_s} \quad (7)$$

$$M_T = M_{end} - M_{start} \quad (8)$$

式中:  $M_{start}$ 、 $M_{end}$  分别为运行的开始累计里程、结束累计里程。

$$M_{TD} = \frac{M_T}{D} \quad (9)$$

$$N_{TD} = \sum_{i=1}^m n_i \quad (10)$$

式中:  $n_i$  为第  $i$  次夜间驾驶的时间长度。夜间驾驶时段考虑为晚上 23 时至次日凌晨 5 时。<sup>[17]</sup>

$$N_{TR} = \frac{N_{TD}}{T_o} \quad (11)$$

$$N_{DF} = \frac{D_{night}}{D} \quad (12)$$

式中:  $D_{night}$  为夜间驾驶的天数。

$$F_{DF} = \frac{D_{fatigue}}{D} \quad (13)$$

式中:  $D_{fatigue}$  为疲劳驾驶天数。疲劳驾驶是指司机驾驶车辆当天驾驶超过 8 h 或连续驾驶超过 4 h<sup>[17]</sup>。

## 2.2 K-Means 算法聚类分析

K-means 算法是一种广泛使用的无监督聚类分析算法,通过将数据集划分为  $K$  个相互独立的子集或簇,在没有先验知识的情况下完成数据的分类。通过选定一组聚类的中心点并最小化每个数据点到其最近中心点的距离,不断迭代直到聚类的中心点不再改变或达到设定的迭代次数<sup>[18]</sup>。

对于包含  $n$  维数据点和要划分为  $K$  个簇的数据集  $X$ ,数据对象之间采用欧氏距离来度量相异性,聚类目标可以表示为:

$$J = \sum_{k=1}^K \sum_{i=1}^{n_k} \|x_i - u_k\|^2 \quad (14)$$

式中:  $K$  是聚类的数量;  $u_k$  是第  $k$  个聚类的中心;  $x_i$  表示数据集中的第  $i$  个点;  $n_k$  是第  $k$  个聚类中的数据点数量。迭代新中心点的计算方法为:

$$u_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i \quad (15)$$

## 2.3 手肘法

本研究利用手肘法判断最佳司机驾驶行为簇数量,其基本思想是通过运行聚类算法并计算不同的簇数量及其对应的误差平方和(sum of the squared

error, SSE)来找到最佳的簇数量。随着  $K$  值的迭代增加, SSE 通常会不断减少。手肘法的意思是找到 SSE 开始急剧减缓的最优  $K$  值。通常,在这个点之后增加更多的簇也无法显著提高模型的性能。

通过绘制 K-SSE 曲线并找到显著拐点,如图 1 所示。由图可见, SSE 变化减弱出现在  $K=3$  时,因此当  $K$  值为 3 时,数据集聚类效果最好。

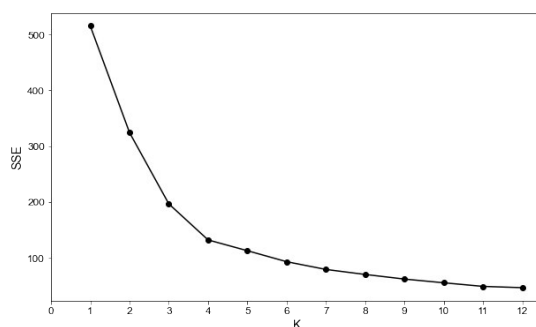


图1 K-SSE 曲线

Fig.1 K-SSE curve

在对司机的驾驶行为进行多维度分析后,本研究成功地将司机行为分为 3 个不同的簇,如表 5 所示。每个簇代表一种特定的驾驶行为模式:

簇 0: 此类别的司机具有较低的总运营时长和运营天数,以及较低的夜间驾驶和疲劳驾驶频率。这可能反映了这一群体的驾驶行为更倾向于白天、短途和非密集运输任务。

簇 1: 此类别的司机展示了更长的总运营时长和运营天数,以及较高的夜间驾驶时长比例和疲劳驾驶频率。这些指标可能揭示了此类司机的运输任务更频繁、更密集,并且可能涉及更多的夜间驾驶。

簇 2: 此类别的司机具有最长的总运营时长和日均运营时长,夜间驾驶时长比例和疲劳驾驶频率也最高。这可能反映了一种高强度、长途的运输模式。

通过构建这些司机运输行为画像,本研究提供了对司机驾驶习惯和运输模式的宏观和微观视角。这些画像不仅涵盖了从驾驶时段到总行驶里程等多个维度,还揭示了与风险相关的统计指标,如夜间驾驶总时长、夜间驾驶频率和疲劳驾驶频率。

## 3 道路拥堵指数抓取

### 3.1 道路拥堵的指标

道路拥堵指数是对道路交通流量和道路容量的一个综合评估指标,旨在描述特定时间和地点的道

表5 司机运输行为特征聚类结果

Tab.5 Driver transportation behavior feature clustering results

Cluster	$T_O$	$T_{OD}$	$T_S$	$T_{SD}$	$E_O$	$M_{start}$	$M_{end}$	$M_T$	$M_{TD}$	$N_{TD}$	$N_{TR}$	$N_{DF}$	$F_{DF}$
0	57.922531	3.041376	18.541667	31.893519	1.813962	2265.666667	125.676371	2.135455	0.047717	0.154631	0.011219	2265.666667	125.676371
1	146.297126	4.966305	29.620690	56.742912	1.898400	6682.068966	225.658111	22.773180	0.194183	0.743399	0.038488	6682.068966	225.658111
2	376.262500	12.542083	30.000000	24.855556	0.828519	2793.000000	93.100000	98.034722	0.258514	0.833333	1.000000	2793.000000	93.100000

路交通状态。百度地图智慧交通采用拥堵指数作为表征交通拥堵程度的客观指标,基于海量的用户定位大数据、交通出行大数据和车辆轨迹大数据等挖掘计算,实时展现各城市和特定路段的交通流动情况。该指数的计算方法主要是通过比对实际行程时间与畅通行程时间,拥堵指数越大代表拥堵程度越高,如表 6 所示。当交通流量增加,导致车速降低时,拥堵指数相应上升,其它如异常天气(如雨、雪、雾等)或者异常道路情况也可能对拥堵指数数值产生影响<sup>[19]</sup>。百度道路拥堵指数为交通规划、城市管理以及驾驶者提供了有价值的参考信息。通过这一指数,可以更好地了解道路交通状况,从而为出行决策、交通导航和城市规划提供数据支持。

表6 道路级拥堵程度划分

Tab.6 Division of road-level congestion index

道路级 拥堵程度	畅通	缓行	拥堵	严重拥堵
划分区间	[1.00~1.50)	[1.50~2.00)	[2.00~4.00)	[4.00~]

3.2 数据抓取

为了获取道路拥堵信息,本研究通过调用百度地图交通出行大数据平台的 API 接口来抓取多源数据。该抓取过程采用了定制的网络爬虫技术,确保数据采集的高效和准确。详细的数据处理流程如下:

- (1) 数据请求:利用 GPS 信息,向 API 接口发送定位请求,从而获取与车辆行驶路径相关的道路拥堵原始数据。
- (2) 数据解析:处理 API 返回的数据,该数据通常采用 JSON 或 XML 格式。进一步从中抽取关键信息,例如:省、市及道路名称,并基于此信息获取特定路段在不同时间的拥堵指数。
- (3) 数据清洗:进行数据的初步处理,包括消除重复数据、修补缺失值以及处理异常值。
- (4) 数据标准化:将所有数据转化为统一格式,使得后续分析和处理更加便捷。
- (5) 异常值处理:通过统计方法检测并处理可能存在的异常值,如突发的交通事故或其他非常态的交通情况导致的数据偏差。

(6) 数据插值:对于数据中可能出现的缺失部分,利用插值方法进行估算,确保数据的完整性。

(7) 数据匹配与融合:数据经过上述处理后,与原始数据集进行匹配,并将拥堵指数作为一个新增列添加到原始数据中,确保数据的完整性和连贯性。

通过上述一系列处理步骤,成功地实现了对 120 辆轻型柴油卡车线路拥堵指数数据的获取。

4 油耗预测模型构建及验证

4.1 模型训练

油耗预测模型结构如图 2 所示,该模型主要分为数据预处理、特征工程、模型构建及训练和结果分析 4 个部分。本研究分别使用 MI 特征筛选方法、司机运输行为 K-Means 聚类算法及网络抓取道路拥堵指数作为模型输入变量,并以 7:3 的比例划分训练集和测试集数据,通过对训练集进行模型训练和参数调试,完成 MLP、RF、XGBoost、MI-Fusion-MLP、MI-Fusion-RF、MI-Fusion-XGBoost 的油耗预测模型建立。最后,利用测试集进行模型检验,并采用多个指标对模型进行评价。

4.2 模型评价指标

本文采用 3 个主要评价指标对模型预测效果进行了评估:平均绝对误差(mean absolute error, MAE)、平均绝对百分比误差(mean absolute percentage error, MAPE)以及拟合优度  $R^2$ 。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

(16)

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

(17)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

(18)

式中:  $y_i$  是第  $i$  个油耗真实值;  $\hat{y}_i$  是第  $i$  个油耗预测值;  $\bar{y}$  是油耗真实值的平均值;  $n$  是样本数量。

模型的预测结果详见表 7,表中概述了各油耗预测模型的精度表现。通过表 7 可以观察到结合考量

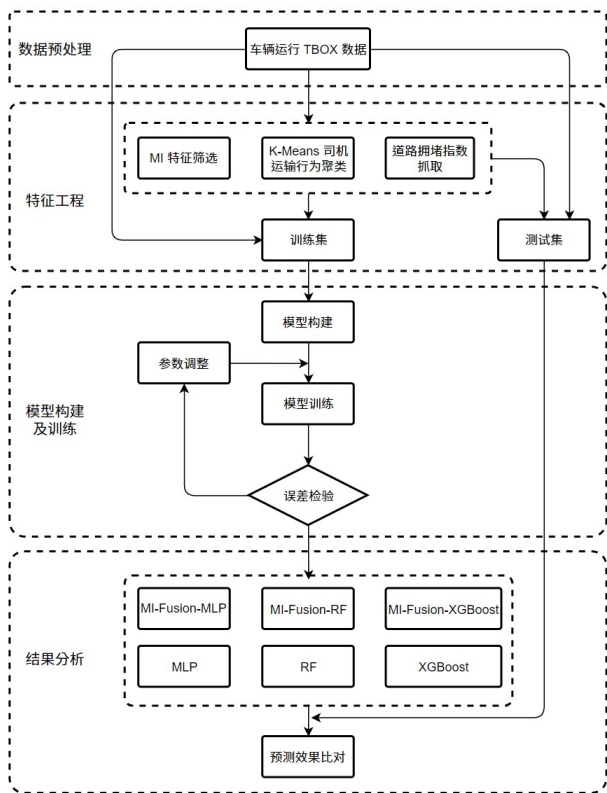


图2 模型训练流程图

Fig.2 Model training workflow diagram

MI特征筛选、司机运输行为以及道路拥堵指数作为输入变量,能够显著提升模型预测的准确性。特别是MI-Fusion-RF模型,其 $R^2$ 、MAE和MAPE分别为0.8715、1.1569和9.46%,与未进行MI-Fusion特征工程的RF模型相比,MAPE提升了3.16%,模型预测效果如图3所示。这一结果充分证明了本文提出的基于MI-Fusion方法构建的油耗预测模型能够有效捕捉车辆运行特征、司机驾驶行为、行驶道路状况与实际燃油消耗之间的映射关系,并证明了其在车辆油耗预测方面的适用性。

### 5 结语

本研究首先使用MI方法筛选了与油耗相关的关键车况特征,并利用实车运行数据构建司机运输

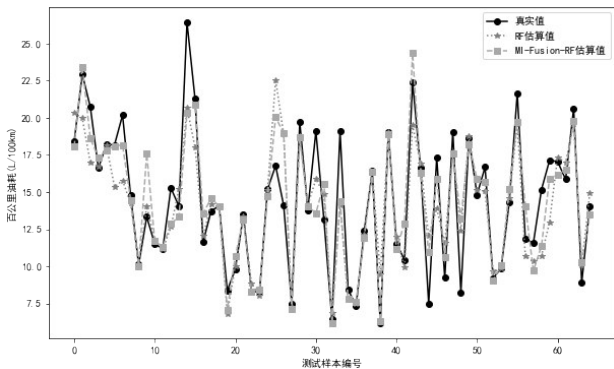


图3 部分模型预测效果

Fig.3 Partial prediction performance of the model

行为画像作为独立特征。然后,基于百度地图交通出行大数据平台的API接口抓取车辆行驶路段道路拥堵指数。进一步,本文结合了XGBoost、RF和MLP机器学习方法,构建基于MI-Fusion特征选择的能耗高精度预测方法。实验结果表明本文提出的预测方法能够显著提高不同驾驶行为和驾驶工况下的能耗预测精度。其中MI-Fusion-RF模型预测准确率达到90.54%,其预测精度相比传统RF模型具有显著提升,预测准确率提高了3.16%。

本研究的成果可为开发预测轻型柴油卡车能耗的通用模型提供实用参考。未来工作可进一步探讨不同类型和规模的车辆数据,考虑车辆载荷、行驶路段坡度、天气等多维因素,寻求更有效的模型设计方法,并在更广泛的应用场景中验证和优化,提高油耗的预测精度和实用性。

### 参考文献:

[1] 国家统计局. 中华人民共和国2022年国民经济和社会发展统计公报[EB/OL]. [2023-02-28]. [http://www.stats.gov.cn/sj/zxfb/202302/t20230228\\_1919011.html](http://www.stats.gov.cn/sj/zxfb/202302/t20230228_1919011.html).  
National Bureau of Statistics of China. Statistical communique on the national economic and social development of the People's Republic of China for the Year 2022[EB/OL]. [2023-02-28]. [http://www.stats.gov.cn/sj/zxfb/202302/t20230228\\_1919011.html](http://www.stats.gov.cn/sj/zxfb/202302/t20230228_1919011.html).  
[2] 中华人民共和国国务院. 两部门联合约谈违规生产货车企业

表7 模型误差验证

Tab.7 Model error validation

评价指标	基本模型			优化模型		
	MLP	RF	XGBoost	MI-Fusion-MLP	MI-Fusion-RF	MI-Fusion-XGBoost
$R^2$	0.725 3	0.761 7	0.728 4	0.847 9	0.871 5	0.850 6
MAE	1.743 5	1.498 4	1.739 1	1.362 1	1.156 9	1.343 8
MAPE/%	14.56	12.62	14.58	11.25	9.46	11.06

- 深入推进货车“大吨小标”专项整治[EB/OL]. (2021-02-07) [2023-02-28]. [https://www.gov.cn/xinwen/2021-02/07/content\\_5585567.htm](https://www.gov.cn/xinwen/2021-02/07/content_5585567.htm).
- The State Council of the People's Republic of China. Two departments jointly hold discussions with truck manufacturers violating regulations to deepen the special rectification of 'overloaded trucks with understated labels'[EB/OL]. (2021-02-07) [2023-02-28]. [https://www.gov.cn/xinwen/2021-02/07/content\\_5585567.htm](https://www.gov.cn/xinwen/2021-02/07/content_5585567.htm).
- [3] WANG H, FU L, ZHOU Y, *et al.* Modelling of the fuel consumption for passenger cars regarding driving characteristics [J]. *Transportation Research Part D: Transport and Environment*, 2008, 13(7), 479.
- [4] YANG X, ZHIZHONG L, MIN-YE C, *et al.* Research on calculation software of fuel consumption for heavy trucks[C]// 2011 Third International Conference on Measuring Technology and Mechatronics Automation. Shanghai: IEEE, 2011, 2: 1121.
- [5] WANG J, RAKHA H A. Fuel consumption model for heavy duty diesel trucks: Model development and testing [J]. *Transportation Research Part D: Transport and Environment*, 2017, 55, 127.
- [6] DU Y, WU J, YANG S, *et al.* Predicting vehicle fuel consumption patterns using floating vehicle data[J]. *Journal of Environmental Sciences*, 2017, 59: 24.
- [7] WYsocki O, DEKA L, ELIZONDO D. Heavy duty vehicle fuel consumption modeling using artificial neural networks [C]// 2019 25th International Conference on Automation and Computing (ICAC). Lancashire: IEEE, 2019: 1.
- [8] 赵晓华,姚莹,伍毅平,等. 基于主成分分析与BP神经网络的驾驶能耗组合预测模型研究[J]. *交通运输系统工程与信息*, 2016, 16(5): 185.
- ZHAO Xiaohua, YAO Ying, WU Yiping, *et al.* Research on a combined driving energy consumption prediction model based on principal component analysis and BP neural networks [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2016, 16(5): 185.
- [9] XU Z, WEI T, EASA S, *et al.* Modeling relationship between truck fuel consumption and driving behavior using data from internet of vehicles [J]. *Computer - Aided Civil and Infrastructure Engineering*, 2018, 33(3): 209.
- [10] WICKRAMANAYAKE S, BANDARA H D. Fuel consumption prediction of fleet vehicles using machine learning: A comparative study [C]// 2016 Moratuwa Engineering Research Conference. Colombo: IEEE, 2016: 90.
- [11] YAO Y, ZHAO X, LIU C, *et al.* Vehicle fuel consumption prediction method based on driving behavior data collected from smartphones [J]. *Journal of Advanced Transportation*, 2020 (4): 9263605.1.
- [12] DU Y, WU J, YANG S, *et al.* Predicting vehicle fuel consumption patterns using floating vehicle data[J]. *Journal of Environmental Sciences*, 2017, 59: 24.
- [13] 黄赫,储江伟,艾曦峰,等. 基于Python的汽车运行油耗预测模型的构建[J]. *电子测量技术*, 2021, 44(20): 113.
- HUANG He, CHU Jiangwei, AI Xifeng, *et al.* Construction of a car fuel consumption prediction model based on Python [J]. *Electronic Measurement Technology*, 2021, 44(20): 113.
- [14] 邹智宏,李超,邓聚才,等. 基于mRMR特征优化算法的油耗预测建模研究[J]. *计算机仿真*, 2023, 40(3): 186.
- ZOU Zhihong, LI Chao, DENG Jucui, *et al.* Research on fuel consumption prediction modeling based on mRMR feature optimization algorithm [J]. *Computer Simulation*, 2023, 40 (3): 186.
- [15] GRANDJEAN E. Fatigue in industry[J]. *Br J Ind Med*, 1979, 36: 175.
- [16] KANEKO T, JOVANIS P. Multiday driving patterns and motor carrier accident risk: A disaggregate analysis [J]. *Accid Anal Prev*, 1992, 24: 437.
- [17] SAGASPE P, TAILLARD J, AKERSTEDT T, *et al.* Extended driving impairs nocturnal driving performances [J]. *PloS One*, 2008, 3(10): e3493.
- [18] YUAN Chunhui, YANG Haitao. Research on K-value selection method of K-means clustering algorithm[J]. *J MDPI*, 2019, 2(2): 226.
- [19] 百度地图智慧交通. 百度地图交通出行大数据平台[EB/OL]. [2023-02-20]. <https://jiaotong.baidu.com/congestion/city/urbanrealtime/>.
- Baidu Maps Intelligent Transportation. Baidu maps traffic and travel big data platform [EB/OL]. [2023-02-20]. <https://jiaotong.baidu.com/congestion/city/urbanrealtime/>.