

# 基于电池能耗的新能源汽车里程预测

房开涛, 李春杰, 朱彦霖

西华大学汽车与交通学院, 四川 成都

收稿日期: 2025年10月23日; 录用日期: 2025年12月30日; 发布日期: 2026年1月7日

## 摘 要

行驶里程作为新能源汽车的一项重要评价指标, 一直备受关注, 对其里程进行高精度预测是当今电池能量管理系统领域的研究热点。本文基于北方某地区车辆1~4月实际行车数据, 划分充放电片段, 对各项参数进行数据清洗、相关性分析, 充分考虑各项电池能耗相关参数特征, 估算出车辆的平均能耗, 进而得出车辆的最大估计里程和最小估计里程, 同时, 结合总电压数据, 我们采用KNN算法对行驶里程进行了回归预测。结果发现, 考虑电池各项参数对新能源汽车里程预测具有较为理想的效果, 通过模型训练测试验证后其均方误差(MSE)稳定在3.6左右,  $R^2$ 达到0.97。具有较高的预测精度。

## 关键词

能量消耗, 行驶里程, 放电片段, KNN

# New Energy Vehicle Mileage Forecast Based on Battery Energy Consumption

Kaitao Fang, Chunjie Li, Yanlin Zhu

School of Automotive and Transportation, Xihua University, Chengdu Sichuan

Received: October 23, 2025; accepted: December 30, 2025; published: January 7, 2026

## Abstract

As an important evaluation index of new energy vehicles, mileage has always been a concern. A study hotspot in the field of battery energy management systems is the high-precision prediction of mileage. Based on the actual driving data of vehicles in a certain area in the north from January to April, this paper divides the charging and discharging segments, performs data cleaning and correlation analysis on various parameters, fully considers the characteristics of various battery energy consumption related parameters, estimates the average energy consumption of vehicles, and then obtains the maximum estimated mileage and minimum estimated mileage of vehicles. At the same

time, combined with the total voltage data, we use the KNN algorithm to predict the mileage. The results show that considering the parameters of the battery has an ideal effect on the mileage prediction of new energy vehicles. After model training and test verification, the mean square error (MSE) is stable at about 3.6, and  $R^2$  reaches 0.97. It has high prediction accuracy.

## Keywords

Energy Consumption, Mileage, Discharge Fragment, KNN

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在双碳战略的发展背景下，能源结构转型加速推进，而新能源汽车也因此取得高速发展。行驶里程作为新能源汽车的一项重要指标，反映着其电池性能以及能量的利用效率，但由于行驶里程受诸多因素的影响，使得实际行驶里程与预计行驶里程总是存在差异，从而严重影响驾乘人员的用车体验，滋生公众的“里程焦虑”，不利于制定合理的行车规划。此外，传统的里程预测方法预测精度较差，难以准确反映其实际行驶里程。因此，结合人工智能和大数据的方法来预测新能源汽车的行驶里程变得愈发重要。本研究聚焦于在车辆行驶过程中，动态地预测其剩余续航里程，旨在为驾驶员提供实时、准确的续航参考。基于此，许多研究学者在近几年的研究中都给予了关注和考证。罗等[1]利用随机森林和 GDBT 两种模型，通过微分思想选择模型的输入和输出，提高预测的准确性，并加以分析比较。胡等[2]考虑环境温度和电池健康状况作为影响行驶里程的主要因素，对其所选择片段进行分层耦合和单点分类来优化预测效果。陈[3]等基于 PSO-RF 算法优选出最佳数据特征组合进行纯电动汽车剩余里程预测。高等[4]将 soc、最小单体电池温度、续航里程等作为多元线性回归模型的关键参数，以电池的荷电状态(soc)和温度为基础，挖掘出潜在的数据特征，共同输入 GDBT 模型，进而提高预测精确度。

综上，目前在新能源汽车里程预测领域的研究多集中在特征参数的选择和高效组合，以及探索更为复杂的预测模型，以此进一步提升预测的精确度。然而，这种过于依赖特定参数和模型优化的方法，往往导致预测结果趋于片面，难以全面、真实地反映车辆的实际运行状况。实际上，新能源汽车的行驶里程深受多种因素影响，其中剩余可用能量和能耗是决定性因素。剩余可用能量直接关联着车辆能够继续行驶的潜力，而能耗则反映了车辆在运行过程中的能量消耗效率。因此，要准确预测新能源汽车的行驶里程，就必须对与能耗相关的各项参数给予充分重视。正是基于这样的认识，本文深入探讨了新能源汽车的能耗特性。通过细致划分放电片段，我们能够更精确地捕捉到车辆在不同运行阶段的能耗变化，进而估算出某段时间内的平均能耗。这一步骤对于理解车辆的能耗模式、预测未来行驶能耗具有重要意义。在此基础上，为实现在每个采样时刻的预测，我们构建了动态特征集：在每个 `daq_time` 时刻，输入特征不仅包含该时刻的实时电池状态参数(如 soc、总电压)，还包含了基于历史放电片段计算得到的近期能耗特征(如前一个完整放电片段的平均能耗)，以及累计行驶里程等，以此反映车辆的实时状态和历史能耗趋势。本文进一步采用了 KNN 算法对行驶里程进行回归预测。KNN 算法作为一种简单而有效的机器学习算法，能够充分利用已知数据点的信息，对未知数据点进行准确预测。通过引入这一算法，能够更加科学地预测新能源汽车的未来行驶里程，为车主提供更加可靠的续航参考。这不仅能够增强新能源汽车的智能化程度，而且对于新能源汽车行业景气度的持续提升也将起到坚实的支撑作用。

## 2. 新能源汽车原始数据预处理

本文所研究的原始数据是以新能源汽车国家大数据联盟所提供的各种实车数据特征，所提供的资料主要是北方某城市在冬后的 2023 年 1~4 月 10 辆同一车型的纯电动汽车实车运行参数，数据包含了汽车信息、外界信息、电池信息等三个方面的内容，包括车速、里程、运行模式、驱动、制动状态、外界环境的最高温度、最低温度值、电池总电压、总电流、电池组最高电压值、最低电压值等，共涵盖了三十多种实际车辆运行数据类型，如表 1 所示，其中汽车内部信息、外部信息、电池信息等三个方面的内容分别是：汽车运行数据、汽车运行信息、汽车行驶数据、车辆运行速度、汽车运营数据。

**Table 1.** Field names and meanings of driving data

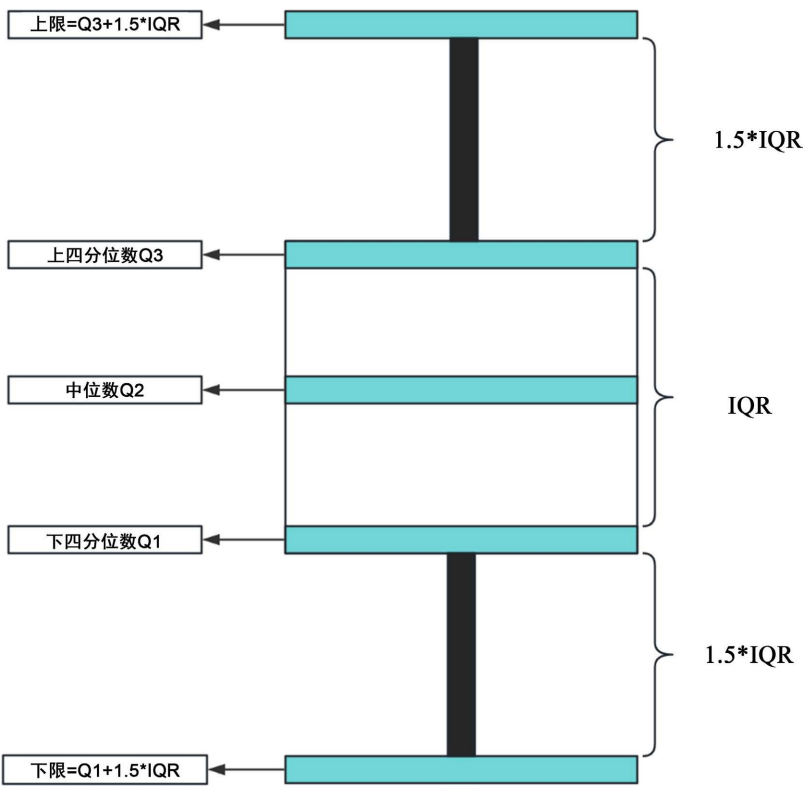
**表 1.** 行驶数据的字段名及意义

字段名	字段意义
vid	在平台上车辆唯一号
daq_time	采样时间
status	车辆状态
c_stat	充电状态
mode	运行模式
speed	车速
mileage	里程
t_volt	总电压
t_current	总电流
soc	soc
dcdc_stat	dc-dc 状态
isulate_r	绝缘电阻
ap_stroke	加速踏板行程值
bp_stroke	制动踏板行程值
gear	挡位
brake_stat	制动状态
power_stat	驱动状态
district	地区
city	城市
province	省份
max_volt_bat_id	最高电压电池子系统号
max_volt_cell_id	最高电压电池单体代号
max_cell_volt	电池单体电压最高值
min_volt_bat_id	最低电压电池子系统号
min_cell_volt_id	最低电压电池单体代号
min_cell_volt	电池单体电压最低值
max_temp_sys_id	最高温度子系统号

续表

max_temp_probe_id	最高温度探针号
max_volt_cell_id	最高电压电池单体代号
max_temp	最高温度值
min_temp_sys_id	最低温度子系统号
min_temp_probe_id	最低温度探针号
min_temp	最低温度值
max_alarm_lvl	最高报警等级
genral_alarm	通用报警标志
motor_nu	驱动电机个数

由于数据收集、存储和处理过程中的误差、数据抓取过程中的噪声、数据生成过程产生的变化等因素，从而使得原始数据出现间歇性缺失，导致出现数据集不完整，数据质量下降等问题，为提高数据质量，减少输入数据自身的不稳定性对机器学习模型精确度的负面影响，对数据进行预处理工作，由于数据本身具有时间序列特点，需按照时间顺序对数据进行排序。此外，由于不可抗力因素的影响导致数据可能存在异常值，缺失值的现象，为避免这些因素对模型的训练和预测产生负面影响，有必要分别采用箱型图法和线性插值的方法对数据中的异常值，缺失值进行处理。箱型图是一种直观、真实的反映原始数据分布情况的数学工具，下**图 1**所示为箱型图的图示说明。



**Figure 1.** Box plot illustration

**图 1.** 箱型图图示说明

图中显示了箱型图的五个典型特征参数,分别为上下限值、上下四分位数和中位数,显著偏离箱型图上下限数据的数据点被认定为异常值,应予以剔除。

由于缺失值可能与其他变量存在潜在的关系,不能随意删除,因此要分别取缺失值左右的邻点对其数值进行估计,用线性插值的方法对缺失值进行填补,图2展示了箱型图法处理异常值(里程)前后变化情况,图3是缺失值处理前后的变化情况,用线性插值的方法对这类数据进行处理,使得缺失值已经被填补。这说明经过处理的数据更能反映此类数据的实际连续状况,以线性插值的方式弥补空缺的数据,以便于后续数据的运行,从而使每一行车过程数据保持完整。

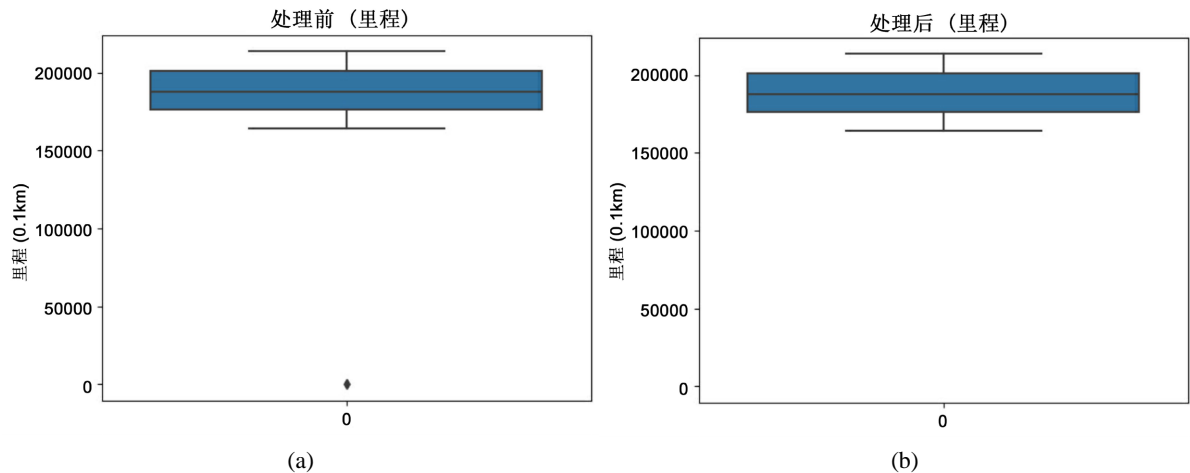


Figure 2. Box plot method for dealing with outliers (mileage) before and after changes

图2. 箱型图法处理异常值(里程)前后变化情况

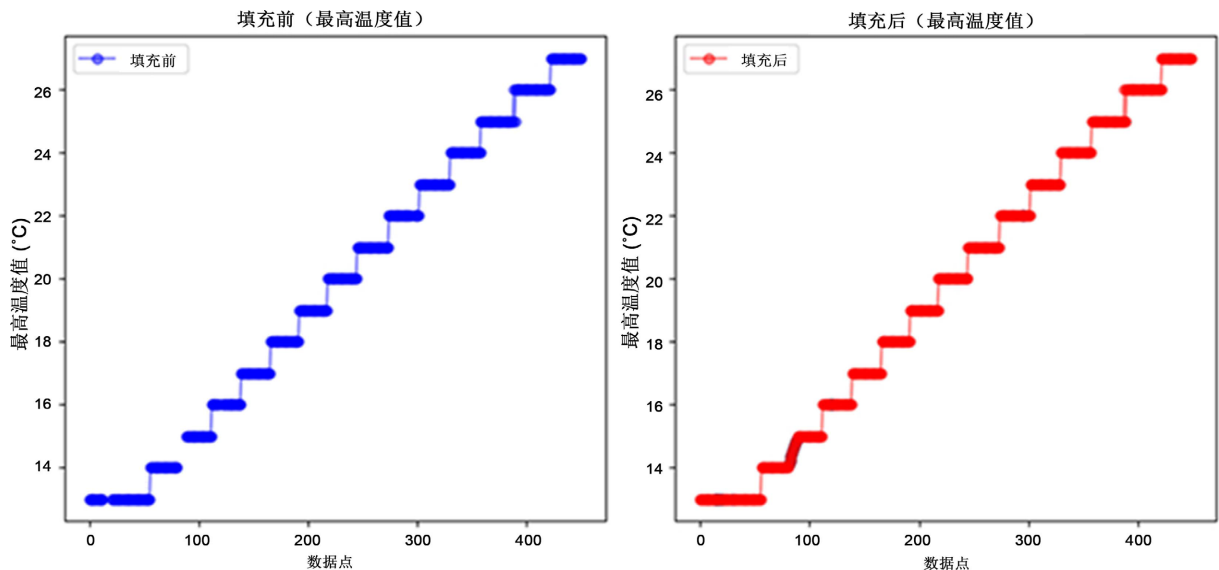


Figure 3. Changes before and after linear interpolation to deal with missing values

图3. 线性插值法处理缺失值前后的变化情况

### 数据特征相关性分析

若将获取到的所有变量均作为训练模型的输入特征,会引起模型可解释性降低,信息冗余[5],伴随

而来的是为训练而耗费的大量时间和计算机资源，导致过拟合现象产生[6]。因此，需要对数据进行选择性的输入，而输入变量的选择则采用 Pearson 相关系数方法进行分析，其公式具体表述如下：

$$r = \frac{\sum_{i=1}^n (x_i - \hat{x})(y_i - \hat{y})}{\sqrt{\sum_{i=1}^n (x_i - \hat{x})^2 \sum_{i=1}^n (y_i - \hat{y})^2}} \quad (1)$$

式中  $n$  为样本数， $x_i$  和  $y_i$  分别为两变量的变量值， $\hat{x}$  和  $\hat{y}$  分别为  $x$ 、 $y$  对应的均值。

通过皮尔逊系数得到采集参数之间各自的相关性，采集参数与目标量(里程)之间的相关性，如图 4 所示，通过热力图呈现各特征之间的相关性，可以直观地反映出图中颜色的深浅程度相关性的强弱，越深的颜色表示相关性越强，可以从图中找到相关的相关性，可以通过 soc、daq\_time、c\_stat、t\_volt、isulate\_r、max\_cell\_volt、min\_cell\_volt 与 mileage 具有较强的相关性根据相关性分析结果表明，影响续航里程的主要因素是电池的各种状态参数和总电压。分别检查数据中 soc 以及总电压为缺失值所在的行并将其删除，遍历数据的每一行，根据充电状态参数来进行数据分割，以每两个相同的充电状态参数行内的 soc 变化量作为一次完整放电，划分出一个完整的放电过程，每个放电过程包含从放电开始到放电结束的所有行为，为确保每个放电过程的数据量足够大，定义每个放电过程中的数据量至少为 300 个。

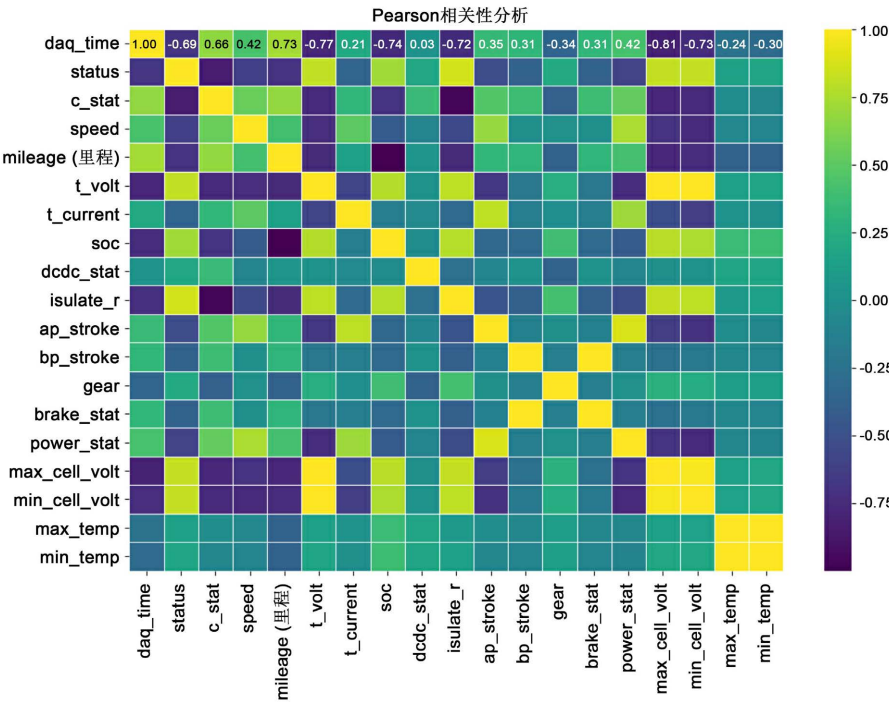


Figure 4. Correlation analysis heat map  
图 4. 相关性分析热力图

最后对每个放电样本计算里程和 soc 的变化值的绝对值，得到累计里程与 soc、累计里程与总电压的相关性，并分别放在两个列表“里程-soc 相关性列表”，“电压-soc 相关性列表”中。对于每个样本，计算从当前 soc 到最终 soc 的变化量，通过 soc 的变化量以及车辆能耗进而得出最大里程估计值和最小里程估计值，以评估电池在不同使用条件下的性能。

此外，为确保数据质量，在前期进行数据清洗工作后，需对电池各项状态参数进行处理和分析，以



此提高预测结果的准确性。根据充电状态和累计里程筛选出特定的数据段(行驶状态), 专注于分析特定条件下的数据, 有助于发现特定规律。计算每个筛选出的数据段的能耗(单位里程的 soc 变化量), 通过计算能耗, 可以评估不同充电周期下的电池性能, 有助于更好地预测剩余里程。图 5 为一段行驶状态下电动车的耗能状态。

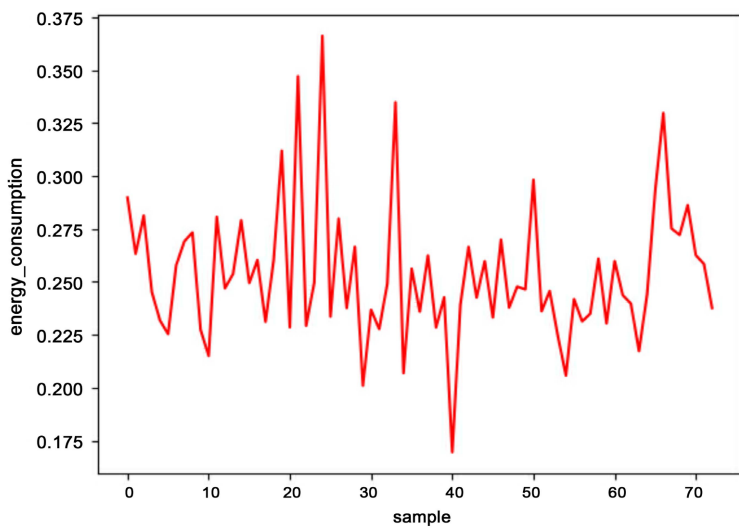


Figure 5. Average energy consumption range  
图 5. 平均能耗范围

图 5 是基于 soc 和总电压所获取的车辆平均能耗范围, 显而易见, 电动汽车的平均功耗大约为 0.25, 其平均功耗在 0.2 到 0.325 的范围之间波动, 这主要取决于新能源汽车在实际行驶过程中所遇到的路况条件。此外, 通过分析能耗的变化规律可以估计平均续驶里程的范围。将平均功耗以及其波动范围代入计算可以得到电动汽车该特定的数据段的真实里程以及对应的最大、最小估计里程。如图 6 所示:

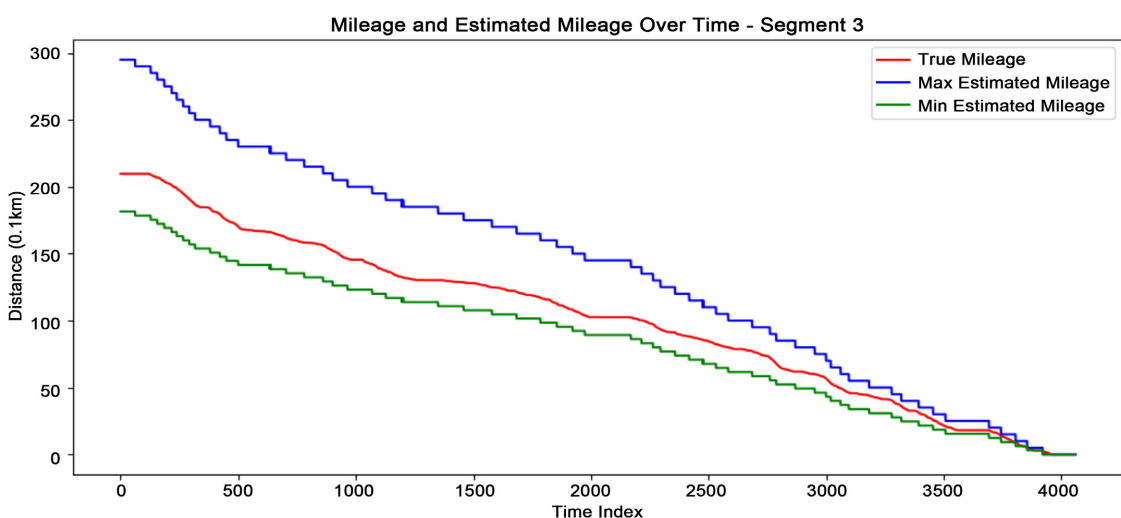
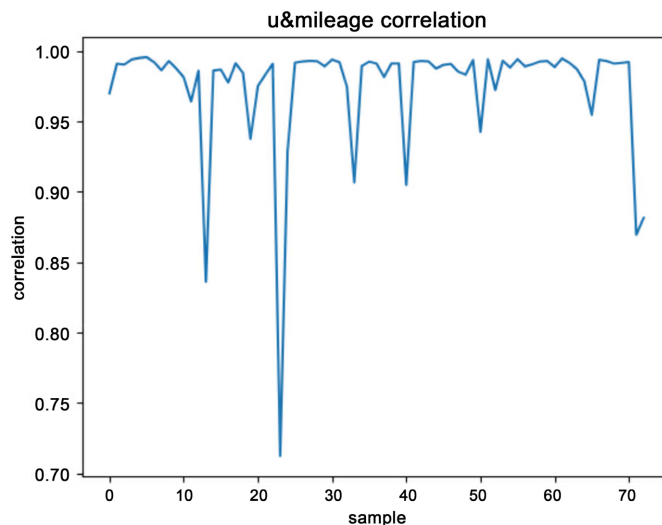


Figure 6. Mileage fitting  
图 6. 里程拟合

由图 6 可知, 在距离电动汽车实际行驶状态更近的最大和最小的估计里程范围内, 电动汽车真实行驶里程始终在这一行驶状态的数据段内波动, 有助于后续分析工作的开展。

进一步通过分析续驶里程和总电压的关系, 得到其相关性如图 7 所示。



**Figure 7.** Correlation between total voltage and cruising range  
**图 7.** 总电压和续航里程的关联度

从图 7 可以明显看出总电压与续航里程存在较强的相关性。后续将在模型输入时作为一个强相关性特征进行输入, 以便得到更为准确的预测结果。

综上所述, 根据电池各项状态参数估算出车辆能耗范围, 以此得出车辆的最大估计和最小估计里程。随后, 将这些数据连同总电压一起作为输入, 利用 K 近邻值算法对电动汽车的续驶里程进行回归分析预测。

### 3. 里程预测模型构建

#### 3.1. KNN 模型基本原理框架

KNN 算法是一种基于实例的学习方法, 由于其独特的性质, 使其既能在分类任务中应用, 又能胜任返回任务的工作。在回归问题中, 它通过计算目标特征与其邻近特征值的平均值来进行预测, 其本质就是预测某个特征的结果由该特征的邻近值决定。此外, 距离度量和 K 值确定是 KNN 算法中两个主要的组成部分, 这也是决定预测精度高低的关键所在。常用的距离度量方式有两种, 一种是欧氏距离, 一种是曼哈顿距离, 如 1 和 2 分别显示其特定的表示方式。

$$A = \sqrt{(x1 - x2)^2 + (y1 - y2)^2} \quad (2)$$

其中  $x1$ 、 $x2$ ,  $y1$ 、 $y2$  分别表示二维空间中的点(特征点),  $x1 - x2$ 、 $y1 - y2$  表示目标特征和已知特征之间的距离。

$$B = |x1 - x2| + |y1 - y2| \quad (3)$$

其中  $x1$ 、 $x2$ ,  $y1$ 、 $y2$  分别表示二维空间中的点(特征点),  $x1 - x2$ 、 $y1 - y2$  表示目标特征和已知特征之间的距离。



K 值的最优选择通常依赖于交叉验证和 Gridsearch 两种方法。具体做法是，将输入模型的数据集划分为训练集和测试集，通过多次训练和测试的过程，来确定能够使预测精度达到最佳的 K 值。

3.2. KNN 模型在行驶里程预测中的应用

本文数据来源于新能源汽车数据大联盟提供的所有实车以及运营数据。包括续航里程，总电压，总电流，电池的荷电状态(soc)，电池单体电压的最高值，以及电池单体电压的最低值，电池最高温度值，电池最低温度值，车速)，将电池相关的各种状态参数和总电压经过数据清洗(移除异常值，补充缺失值)后作为模型输入，一起输入到 KNN 模型中。为评估模型的时序泛化能力，我们采用了严格的时间序列分割方案：使用 2023 年 1 月至 3 月的数据作为训练集，2023 年 4 月的数据作为测试集。此外，我们还引入了一个外部验证集，该数据集来源于同车但在不同年份(2024 年同期)采集的车辆数据，以进一步检验模型在不同时间段和潜在不同使用条件下的泛化性能。然后对模型进行训练测试集划分(训练集中数据量占 80%，测试集中数据量占 20%)，根据数据训练结果判断是否需要交叉验证，网格搜索寻找调整 K 值，从而提高预测准确度。最后通过相同车辆在其他年份的数据验证经训练测试后的模型精度。具体流程如图 8 所示。

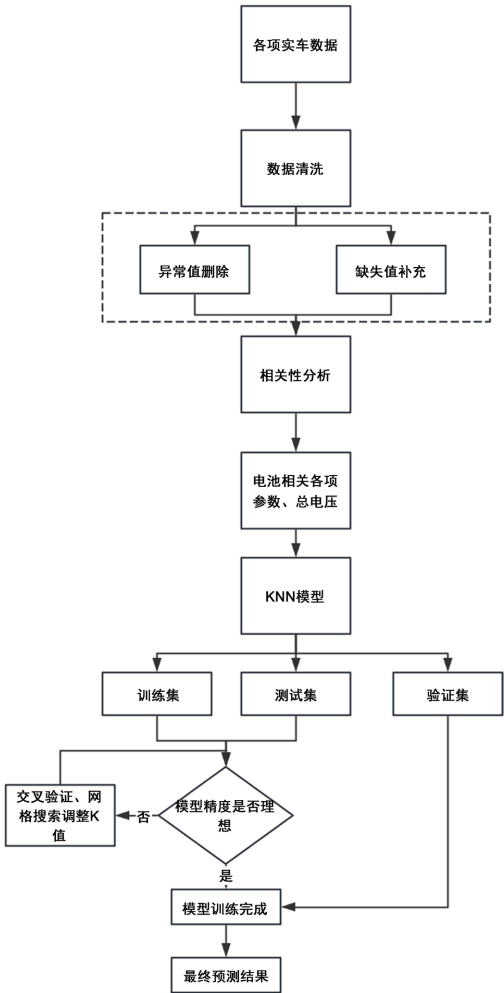


Figure 8. KNN model mileage application process  
图 8. KNN 模型行驶里程应用流程

### 3.3. KNN 模型参数设置

为了提高 KNN 模型的预测精度、实用性和兼容性,需对模型各项参数进行调整设置,经过多次试验测试,设置 KNN 模型邻近点个数为 3,采用曼哈顿距离和欧式距离的倒数作为权重模型精度效果较好,后续多轮训练测试验证发现上述各项模型指标适用于经预处理后的数据集,具有较高的匹配性,对模型精度起着有益作用。

### 3.4. KNN 模型预测结果

在各种新能源汽车电池的相关特征参数、总电压输入数据清洗后的 KNN 模型中,我们根据上文提到的车型调校和测试集进行比例划分,对数据进行相应的划分,然后再进行车型调校,用测试集来验证车型的精确性,从而确保续航里程预测结果的可靠性,我们在此基础上,根据上文提到的车型调校和测试鉴于预测精度的测评指标众多,如平方误差、均方误差、平方根误差、平均绝对误差、决定系数以及校正决定系数等,在行驶里程预测精度的衡量上,本文特别采用均方误差(MSE)和决定系数( $R^2$ )两个指标作为衡量标准。其中, MSE 值越小说明预测的准确性越高,而  $r$  值与 1 的距离越近,则说明模型的拟合效果越好。均方误差和决定系数的具体计算方式分别在式(4)和式(5)中给出。

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

- $n$ : 样本总数,确保误差平均化。
- $y_i$ : 第  $i$  个样本的真实值(实际观测值)。
- $\hat{y}_i$ : 第  $i$  个样本的预测值(模型输出值)。
- $(y_i - \hat{y}_i)^2$ : 单个样本的误差平方,通过平方运算放大较大误差的影响。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

其中,分子部分  $\sum_{i=1}^n (y_i - \hat{y}_i)^2$  真实值与预测值的平方差和(即误差平方和),而分母部分  $\sum_{i=1}^n (y_i - \bar{y})^2$  则表示真实值与其均值平方差的总和(即总平方和)。 $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  为真实值的均值。

本文所采用的新能源汽车行驶里程预测评价指标与式(4)和式(5)的计算原理一致,即通过已有的行驶里程和预测的行驶里程来评价模型精度。

图 9、图 10 和图 11 分别展示了行驶里程预测模型在训练集、测试集以及验证集上的预测效果。

结合图 9、图 10、图 11 和表 2 可以看出,根据其实际行驶情况,车辆实际续驶里程一般在 0~800 公里范围内浮动。此外,从图 9、图 10、图 11 还可以看出,无论是训练集测试集,还是验证集,预测续驶里程和实际续驶里程的拟合效果非常理想,其均方误差(MSE)和决定系数( $R^2$ )也体现出模型有着较高的预测精度,在训练集、测试集和验证集上,该模型的均方误差(MSE)分别为 3.60、3.72 和 3.95;决定系数( $R^2$ )在不同数据集上的结果分别为 0.974、0.962 和 0.945。综上,该模型具有较强的数据集适应性,能够准确地预测新能源汽车的续驶里程,考虑电池各项相关特征参数和总电压作为主要模型输入数据,可以使模型表现出非常理想的精度和效果,对于缓解驾乘人员因续航里程而产生的焦虑情绪以及优化电动汽车的充电策略具有非常重要的意义。

## 4. 总结展望

本文基于 KNN 算法对行驶里程进行预测,通过对数据进行分析处理,发现电池各项状态参数和总电压是影响续驶里程的主要原因。基于线性关系,构建包含电池、总电压和续航里程等各种状态参数的多

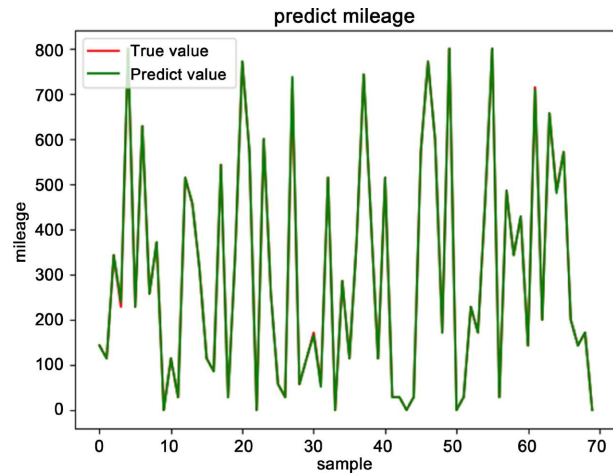


Figure 9. Training set model results

图 9. 训练集模型结果

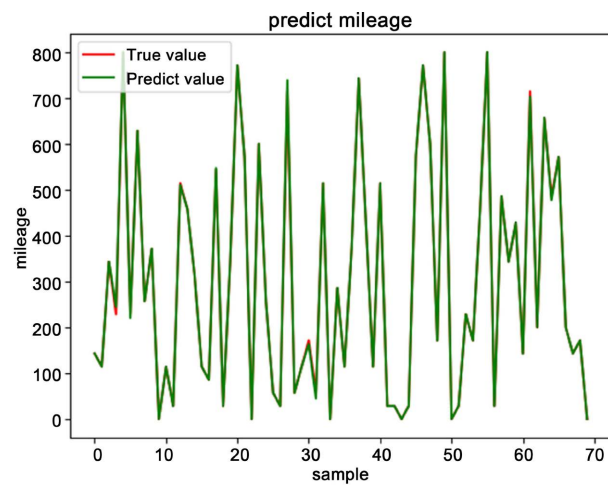


Figure 10. Test set model results

图 10. 测试集模型结果

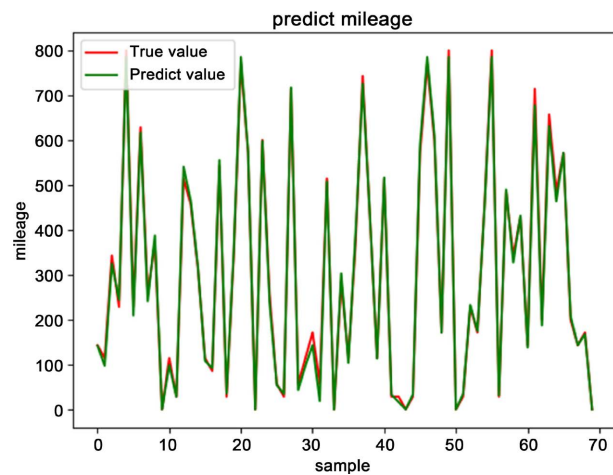


Figure 11. Validation set model results

图 11. 验证集模型结果

**Table 2.** Accuracy evaluation indicators on the training, testing and validation sets  
**表 2.** 训练测试验证集上的精确度考核指标

数据集	训练集	测试集	验证集
MSE	3.60	3.72	3.95
R <sup>2</sup>	0.974	0.962	0.945

元素线性回归模型。为了进一步提升模型的准确性，本文将线性模型计算所得的残差作为一个新增特征引入。经过这一改进，KNN 回归预测模型展现出了更高的精度，有效地弥补了多元线性回归模型存在的不足。

因此，考虑电池各项状态参数和总电压有着较为理想的续驶里程预测效果，解决了传统预测方法预测精度差、适应性差的难题，进一步优化了预测效果，解决模型泛化问题，证明了 KNN 算法用于里程预测的可行性。本研究模型表现优异的主要原因在于：1) 通过划分充放电片段并计算平均能耗，有效捕捉了车辆在不同阶段的能量消耗模式，提供了关键的能耗趋势信息；2) 结合实时电池状态(如 soc、总电压)与历史能耗特征，构建了能够反映车辆动态运行状况的综合输入特征集。然而，本研究也存在一些局限性。首先，KNN 算法作为一种惰性学习方法，在预测时需要存储全部训练数据并进行实时距离计算，可能导致在线预测时计算和存储开销较大。其次，模型性能可能对驾驶行为的突然改变(如急加速、急减速等剧烈驾驶风格变化)以及未在训练数据中充分体现的极端环境条件(如极寒天气)较为敏感。此外，模型输入特征主要基于车辆自身状态和电池参数，尚未充分考虑外部交通环境(如实时路况拥堵程度)对能耗的直接影响。此外，为了进一步提升预测精度以及更加真实地反映新能源汽车在不同驾驶状态、不同驾驶路段的续驶里程，还可以进行以下工作：

- 1) 可以考虑驾驶员的驾驶风格和习惯，并结合所在地区的实际环境来综合评判。
- 2) 进一步分析放电片段的数据特征，划分出更为细小的行车工况(城区、郊区、高速等)。
- 3) 结合采集的实车数据具有时序特点，建立更为复杂的预测模型。

### 参考文献

- [1] 罗红梅, 周逸凡. 基于大数据分析的电动汽车行驶里程预测方法研究[J]. 汽车实用技术, 2020(13): 1-4.
- [2] 胡杰, 翁灵隆, 覃雄臻, 等. 基于多模型融合的电动汽车行驶里程预测[J]. 交通运输系统工程与信息, 2020, 20(5): 100-106+141.
- [3] 陈德海, 任永昌, 华铭, 等. 基于 STM32-OCV 法的纯电动汽车剩余里程预测[J]. 电子技术应用, 2017, 43(12): 33-35+39.
- [4] 高航. 基于机器学习的纯电动汽车的行驶里程预测研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2018.
- [5] 张二伟, 肖文书. 一种基于相关系数的距离估计算法[J]. 现代雷达, 2014, 36(11): 58-61.
- [6] 李俭川, 秦国军, 温熙森, 等. 神经网络学习算法的过拟合问题及解决方法[J]. 振动、测试与诊断, 2002(4): 16-20+76.