

基于用户数据的云端用户失效风险等级辨识

姚 涛

上海理工大学机械工程学院，上海

收稿日期：2024年12月25日；录用日期：2025年1月18日；发布日期：2025年1月24日

摘要

相较于传统汽车，电动汽车的驱动电机展现出更宽广的调速范围、更大的启动扭矩、更高的功率密度及效率。然而，随着转速、扭矩等载荷强度的提升，电驱动系统面临更高的失效风险。随着新能源汽车的不断发展，市面上有大量用户，开展不同用户等级同一部件损伤风险评估，开展同一用户等级不同部件损伤风险评估，揭示用户等级与电驱损伤关联规律。结合聚类后用户风险等级与部件损伤关联结果，设计以用户数据核心部件损伤为输入、用户风险等级为输出的机器学习模型，实现云端新增用户的失效风险等级快速辨识。

关键词

用户大数据，电驱动系统，机器学习，风险等级

Identification of Cloud User Failure Risk Level Based on User Data

Tao Yao

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Dec. 25th, 2024; accepted: Jan. 18th, 2025; published: Jan. 24th, 2025

Abstract

Compared with traditional vehicles, the drive motor of electric vehicles shows a wider speed range, larger starting torque, higher power density and higher efficiency. However, with the increase of load strength such as speed and torque, the electric drive system faces a higher risk of failure. With the continuous development of new energy vehicles, there are a large number of users on the market to carry out damage risk assessment of the same component at different user levels, and carry out damage risk assessment of different components at the same user level, so as to reveal the correlation law between user level and electric drive damage. Combined with the results of the association

between the user risk level and the component damage after clustering, a machine learning model with the damage of the core component of the user data as the input and the user risk level as the output is designed to realize the rapid identification of the failure risk level of the new users in the cloud.

Keywords

User Big Data, Electric Drive System, Machine Learning, Risk Level

Copyright © 2025 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

随着当前汽车行业的迅猛发展与汽车保有量的显著提升[1]，其对能源资源的消耗及环境质量的负面影响日益加剧。鉴于此，汽车行业正逐步转型，致力于新能源汽车领域的深入探索。国内外汽车制造商纷纷投身于纯电动汽车、混合动力汽车以及氢燃料电池汽车等多种类型新能源汽车的研发之中。近年来，电动汽车总体市场占有率目前较高，赵礼辉等[2]提出了车辆使用年限主要集中在6~12年之间，车辆年行驶里程与期望使用年限间存在较好的幂函数关系，期望使用年限随年行驶里程的增加而减小，95%用户全寿命周期行驶里程小于 70×10^4 km。

新能源汽车与传统燃油车相比，驱动电机系统代替了燃油机，使得新能源汽车具有清洁污染、效率高、噪音小等优点。此外，新能源汽车的动力由电驱动系统提供，电驱动系统中电机和减速器的集成化使得其体积更小、效率更高、运行平稳。

电驱动系统主要是由驱动电机、电机控制器与减速器这三大子系统组成的。对于电驱动系统驱动方式主要包括分布式驱动和集中式驱动。针对电驱动系统各部件的定义可以依据 GB/T 18488.1-2015 标准的介绍[3]。电驱动系统实际上是复杂的机电一体化系统，包含不同的零部件，不同的零部件具有不同的失效模式，关联多种物理载荷特征。对于电驱动系统可靠性的研究，首先需要明确各部件的失效机理，这一标准[4]旨在通过全面的可靠性试验方法，为电动汽车制造商提供系统的测试依据，确保电动汽车驱动电机系统的质量和可靠性，从而提高消费者对电动汽车的信任度。

结合电驱动系统零部件失效机理构造表征用户风险程度的特征参数，采用多维特征融合方法进行参数降维，建立合理的目标函数，然后采用非监督学习算法实现典型用户聚类，分析不同类别的用户与不同部件的关联，建立失效风险综合评价指标，开展不同用户等级同一部件损伤风险评估，开展同一用户等级不同部件损伤风险评估，定义用户损伤等级。

基于用户原始载荷时域数据，通过计算电驱动系统不同零部件的伪损伤，进而计算出不同零部件的损伤强度，结合聚类后用户风险等级与部件损伤关联结果，设计以损伤强度信号为输入，用户风险等级为输出的机器学习模型，实现云端新增用户的失效风险等级快速辨识。

2. 电驱动系统核心部件失效机理

电驱动系统中不同零部件的失效模式和失效机理也不相同，针对不同的零部件，有不同的受迷宫模型，针对不同的寿命损伤模型，构建不同的伪损伤计算方法，各核心部件其受到的载荷一方面承受汽车在长期服役过程中路面传递的激励载荷[5]，另一方面承受电驱动系统在运行过程中各部件受到多物理场

(力、热、电、磁)载荷的影响[6]。对于用户来说,不同用户的驾驶习惯也是完全不同的,即使地域环境与道路工况类型是相同的情况下,其载荷特征与损伤效果也会具有很大的差异性。因此这也会对电驱动系统可靠性提出了更高的要求,基于实际用户载荷分析电驱动系统各部件失效载荷与失效机理可为电驱动系统零部件设计提供参考依据,并指导产品可靠性试验验证。

电驱动系统中包含多种部件,各部件关联载荷也是不同的[7],并且不同部件的失效模式和失效机理也不相同。对于电驱动系统失效部件来说,主要包括机械结构部件(轴、齿轮、轴承、磁轭、壳体等);电子电气部件(控制单元、功率器件、电容、逆变器、绕组等)。对于轴类部件来说,失效模式通常为机械疲劳失效,其失效主导载荷来源于扭矩的变幅加载历程[8]。对于齿轮部件,其失效主要为接触疲劳损伤与弯曲疲劳损伤,在用户运行工况下,不同扭矩之间的交变也会对齿轮造成损伤,而转速的不同,也会影响着齿轮转动的圈数。对于轴承类部件来说,失效模式主要为接触疲劳,当轴承要在高速运行工况下进行运行的时候,会引起温度升高,进而加速轴承失效。

目前许多学者经过大量试验与对不同部件的损伤研究,针对不同的部件建立不同的寿命预测模型,最为典型的是 Palmgren-Miner 准则[9]。假设材料发生疲劳失效时[10],所吸收的能量为 W,在不同载荷等级 S_i 作用到失效时的循环数为 N_i ,载荷 S_i 作用 n_i 次时材料所吸收的能量 W_i ,那么对于不同的载荷等级 i ,有:

$$\frac{W_i}{W} = \frac{n_i}{N} \quad (1)$$

当吸收的能量值达到极限时,材料发生失效,则有:

$$D = \sum_{i=1}^n \frac{n_i}{N_i} = \sum_{i=1}^n \frac{W_i}{W} = 1 \quad (2)$$

式(2)中, n 个载荷等级下总的疲劳损伤为 D ,某一载荷下造成的疲劳损伤 D_i 与载荷循环比成线性关系。

在工程应用上,对于不同载荷谱作用于同一零部件时,采用伪损伤的方法可以来反映不同载荷谱对零件的损伤效果。伪损伤计算是以应力与寿命的曲线为基础,利用 miner 损伤准则,计算在不同载荷等级下所对应的损伤,进行累加损伤。S-N 曲线的表达式为:

$$N = \alpha \cdot S^{-\beta} \quad (3)$$

式(3)中, S 为应力幅值, N 为试件在幅值 S 作用下的疲劳寿命, α 为常数、与材料有关, β 为疲劳强度指数。

3. 风险用户等级定义

针对电驱动系统中各零部件的复杂失效机理与多样化的失效模式,本研究构建了一套详尽的特征参数体系,旨在精确量化并表征用户所面临的风险级别。为了显著提升数据处理效率并优化模型预测精度,我们引入了多维特征融合技术与主成分分析(PCA)方法[11],对原始高维数据进行有效降维处理。在此基础上,我们设计了合理的目标函数,并采纳了非监督学习领域的经典算法——Kmeans 聚类算法,对用户群体进行了深入的聚类分析,以期实现典型用户特征的精准划分。

基于聚类分析的结果,我们进一步对不同等级的用户群体进行了细致的剖析。在同一等级的用户类群内部,我们评估了不同零部件的失效风险等级,以揭示其内部差异与共性特征。同时,我们也对比分析了不同等级用户在同一部件上的失效风险等级,以探究风险等级随用户群体变化的规律。

3.1. 特征参数构造

在探究风险用户等级划分的深层次过程中,特征参数的遴选与用户驾驶行为的关联性显得尤为重要。

驾驶行为的多样性，涵盖操作习惯、行驶路径的选择偏好以及车辆维护保养的重视程度等多个维度，均会对机械部件的损耗程度产生直接或间接的影响。本研究致力于构建一个既科学又全面的特征参数体系框架，旨在实现对用户风险等级更为精准的量化评估。具体而言，我们选取了零部件轴的磨损状况、齿轮的损伤程度以及轴承的损坏情况作为核心特征参数，这些参数深刻而全面地映射出电驱动系统的整体损伤面貌，如表 1 所示，展示了部分用户的关于各部件的损伤。

Table 1. Feature parameter construction**表 1.** 特征参数构造

轴	齿轮	轴承
5.4E-07	8.8E-10	2.3E-08
1.2E-06	3.6E-09	3.9E-08
4.8E-07	6.4E-10	2.6E-08
4.7E-07	9.6E-10	2.4E-08
6.9E-07	1.6E-09	3.2E-08
5.3E-07	1.0E-09	2.5E-08
6.1E-07	9.8E-10	2.6E-08

3.2. 主成分分析

成分分析(PCA)是一种广泛应用的降维方法，它通过线性变换将原始的高维数据映射到一个低维空间中，同时尽可能保留数据中的主要变化模式。每个主成分都是原始变量的线性组合，它们按照方差大小进行排序，每一个主成分都是由原始变量的线性组合得到的结果，当我们通过选择前几个主成分的时候，我们可以在降低数据维度的同时也能保留大部分的有效信息。

(一) 主成分基本原理

主成分基本原理的基于数据的方差最大化思想。方差代表了数据的离散程度，方差越大，说明数据在该维度上的变化越丰富，所包含的信息也就越多[12]。因此，PCA 通过找到一个正交变换矩阵，将原始数据投影到一个新的坐标系中，使得投影后的数据在每个维度上的方差尽可能大。这些新的维度就是主成分，它们之间互不相关，且能够尽可能多地保留原始数据的信息。

具体来说，PCA 的实现过程包括以下几个步骤：

数据标准化：这一步是将原始数据进行标准化的处理，当标准化以后，原始的数据都会在 0~1 之间。这将会消除量纲和数值大小对分析结果的影响。

计算协方差矩阵：计算标准化后数据的协方差矩阵，反映各变量之间的相关性。

特征值分解：将计算得到的协方差矩阵进行特征值分解，分解后便会得到特征值和特征向量。

选择主成分：根据特征值的大小选择前 k 个主成分。

转换数据：将原始数据转换到由前 k 个主成分构成的新坐标系中，得到降维后的数据。

(二) 主成分分析结果

将本文构建所有的用户数据，依据轴、齿轮、轴承这三个核心部件的失效机理，与寿命模型，分别计算轴、齿轮、轴承的损伤，这三维特征参数进行主成分分析，依据累积信息贡献率大于 90% 提取出前两个主成分，如表 2 所示，该表展示了部分用户的主成分分析结果，在该表中，每一列代表一维主成分，每一行则代表一个用户。

Table 2. Principal component analysis results
表 2. 主成分分析结果

主成分一	主成分二
-1.17	0.36
0.66	-1.05
0.81	0.84
-0.37	-0.17
-0.45	-0.20
-0.16	1.26
-0.37	-0.11
-0.60	0.66
0.28	0.18

3.3. 聚类分析

本文采用非监督学习领域的 K-means 聚类算法，旨在不明确预设用户分类标准的情况下，对用户群体进行典型用户划分。具体而言，我们首先设定聚类数目 k 值，即预设将用户划分为 k 类。在算法初始阶段，随机选定 k 个初始类簇中心。随后，依据点与点之间的欧氏距离(或曼哈顿距离、马氏距离)，将每个用户(即数据点)分配到距离最近的类簇中心所代表的类别中。这一过程不断迭代进行，直至类簇中心点的位置变化趋于稳定，或达到预设的迭代次数上限。整个 K-means 聚类算法的具体实施流程如图 1 所示，通过此流程，我们能够有效地将用户群体划分为 k 个具有显著特征的典型用户类别。

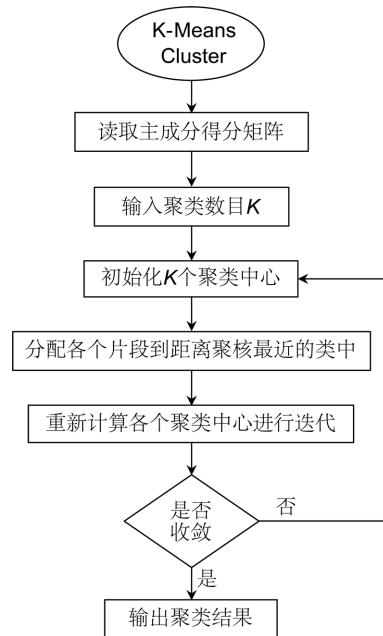


Figure 1. K-means algorithm flow
图 1. K-means 算法流程

聚类中输入的样本数据为降维后的主成分得分矩阵，假设样本数据为矩阵 $X_{n \times m}$ ，其中 n 为样本个数， m 为对应的特征维度。初始化 k 个聚类中心 $\{C_1, C_2, \dots, C_k\}$ ，计算每个样本到每个聚类中心的欧式距离，

其公式如下：

$$dis(X_i, C_j) = \sqrt{\sum_{t=1}^m (X_{it} - C_{jt})^2} \quad (4)$$

式(4)中， X_i 表示第 i 个样本， $1 \leq i \leq n$ ； C_j 表示第 j 个聚类中心， $1 \leq j \leq k$ ； X_{it} 代表第 i 个样本的第 t 个特征， $1 \leq t \leq m$ ； C_{jt} 代表第 j 个聚类中心的第 t 个属性。

将所有的距离通过算法得出后，通过比较每个样本到每个聚类中心之间的距离，进而将样本分配到距离其最近的聚类中心的簇中，便得到 k 个类簇。

3.4. 聚类数目确定

对于聚类数目 k 的选取，肘部法则(Elbow method)是随着 k 值增大时，计算损失函数下降的拐点，其损伤函数为簇内误差平方和(SSE)，簇内的样本点到聚类中心的距离反映了簇内的凝聚度，则对于簇内误差平方和计算公式如下：

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (5)$$

式(5)中， C_i 是为第 i 类簇， p 是 C_i 中的样本点， m_i 是 C_i 的聚类中心点。

当聚类数目为 2 到 7 时，分别计算误差平方和 SSE 结果如图 2 所示。当聚类数目为 5 时，近似位于曲线拐点，原因是当聚类数目小于 5 时，损失函数 SSE 下降速度较快，各类别之间容易区分；当聚类数目大于 5 时，损伤函数 SSE 下降速度较慢，即使增加一个类别，SSE 值下降很小，各类簇间不易区分。因此取聚类数目为 5。

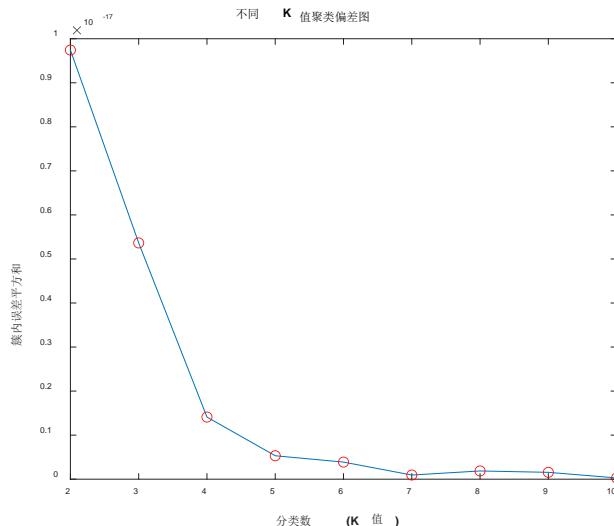


Figure 2. Error squared and SSE
图 2. 误差平方和 SSE

3.5. 聚类结果分析

通过 K-means 聚类分析以后，3703 名用户被分为五类，将每类的用户数目进行统计，等级一有 220 名用户，等级二有 410 名用户，等级三有 780 名用户，等级四有 532 名用户，等级五有 1761 名用户，结果如表 3 所示。

Table 3. Number of users result
表 3. 用户数目结果

等级一	等级二	等级三	等级四	等级五
220	410	780	532	1761

3.6. 同一等级类内不同部件失效风险等级评估

通过开展对同一用户等级类内不同部件失效风险等级评估，可以得到每类用户与各部件失效风险相关性，如图 3 所示，等级一的用户主要为齿轮的失效风险等级较高，等级二的用户主要为轴的失效风险等级较高，等级三的用户主要为轴承、齿轮的失效风险等级较高，等级四的用户对各个部件的损伤差异都比较小，等级五的用户对轴与轴承的损伤都比较大。

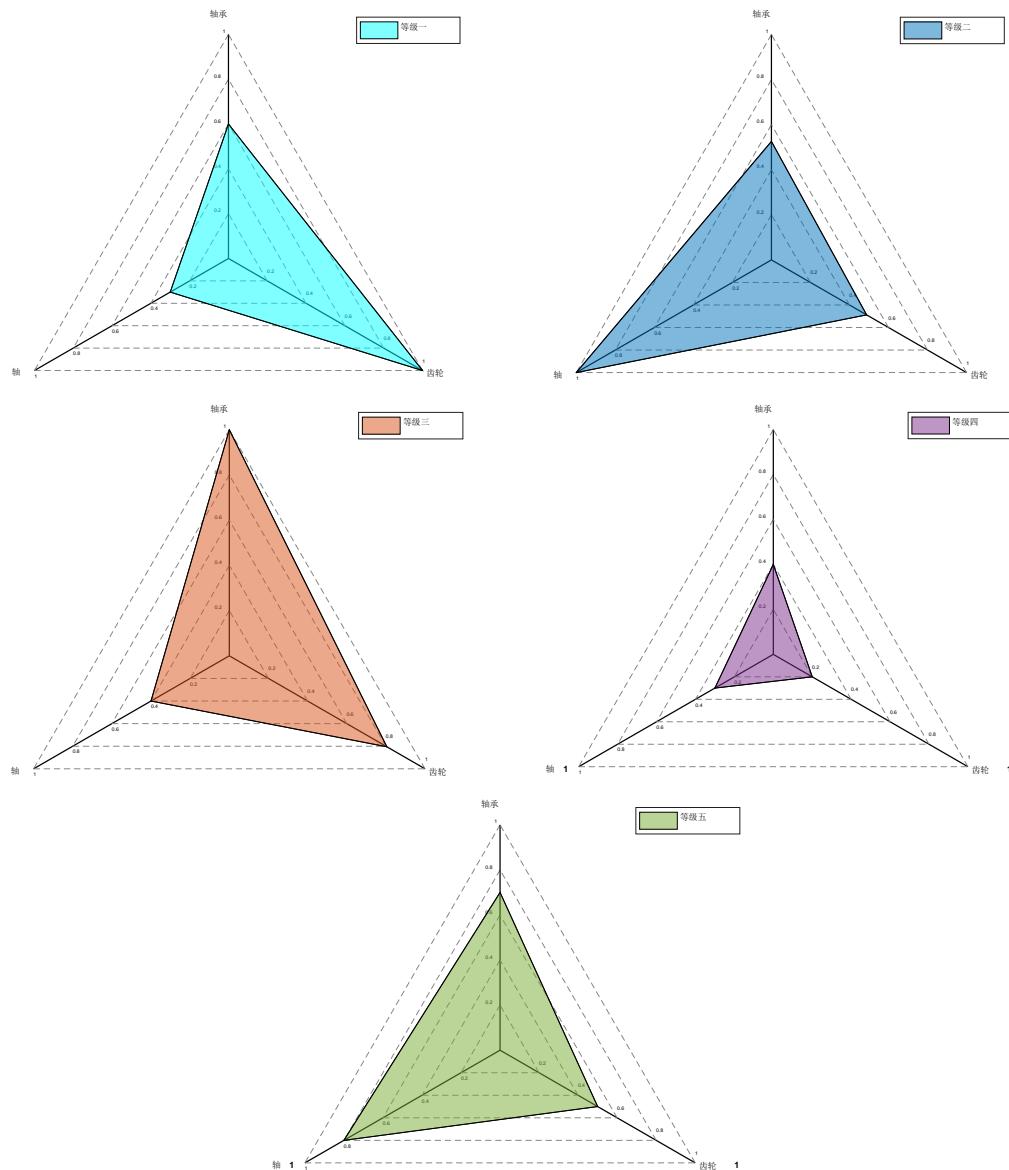


Figure 3. Assessment of the failure risk level of different components within the same class
图 3. 同一等级类内不同部件失效风险等级评估

3.7. 不同等级同一部件失效风险等级评估

通过开展对不同等级用户同一部件失效风险等级评估，可以揭示各部件与各等级用户之间的失效风险相关性，如图 4 所示，等级一的用户主要为轴承失效风险等级最高，等级三的用户主要为轴、齿轮、绕组、IGBT 的失效风险等级最高。

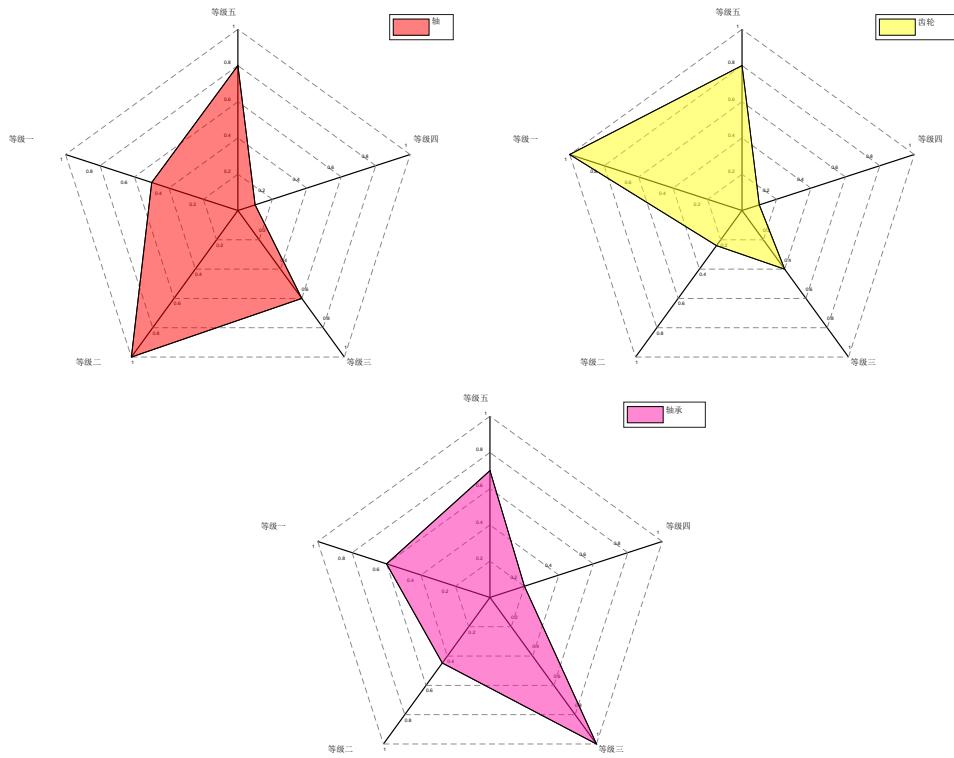


Figure 4. Assessment of the failure risk level of the same component at different levels
图 4. 不同等级同一部件失效风险等级评估

4. 云端用户失效风险等级辨识

4.1. 随机森林算法模型

在机器学习中，随机森林作为一种监督式学习算法，广泛应用于分类和回归问题。它以决策树为基础进行构建，并通过集成多个决策树来形成一个强大的模型。随机森林并不依赖于整个训练数据集来构建单一的、大型的决策树。相反，它采用多个小的决策树，这些决策树是基于训练数据的不同子集和随机选择的特征属性来构建的。随后，随机森林将这些决策树的结果进行组合，以形成一个更为准确和稳健的模型。随机森林中的每个子集都是通过随机抽样的样本和随机选择的特征来创建的。这种随机性降低了决策树对训练数据的过度依赖，有助于减少过拟合的风险。通过这种方式，随机森林能够提升模型的泛化能力，使其在处理新数据时表现更为出色。

在决策树中，每个节点表示一个特征的测试，每个分支表示一个测试结果。决策树的目标是通过一系列的特征测试，将数据分成尽可能纯的子集。这个过程是通过递归地选择最优特征进行分割来实现的，最优特征的选择通常基于某种度量标准，如信息增益，信息增益越大，数据集的纯度提升的越多，不确定性减少的越多，信息熵用于度量一个随机变量不确定性的大小，是度量信息含量的一个指标。计算公式如下

$$Hx = -\sum_{i=1}^{i=n} p_i \log_2(p_i) \quad (6)$$

式中, x 是数据集, n 是数据集中的类别数, p_i 是第 i 个类别在数据集中出现的概率。

4.2. 用户特征参数输入

风险用户等级定义中有五类用户, 将五类用户输入随机森林算法模型进行训练, 如表 4 所示, 等级一共 220 名用户, 等级二共 410 名用户, 等级三共 780 名用户, 等级四共 532 名用户, 等级五共 1761 名用户, 一共 3703 名用户。

Table 4. Number of users result

表 4. 用户数目结果

等级一	等级二	等级三	等级四	等级五
220	410	780	532	1761

4.3. 用户风险等级辨识结果

3703 名用户核心部件损伤作为输入, 5 个用户类别作为输出, 其中 3000 名用户作为训练集, 703 名用户作为测试集, 如图 5 所示, 预测集的准确率达到 100%, 如图 6 所示, 测试集的准确率达到 96.9697%。

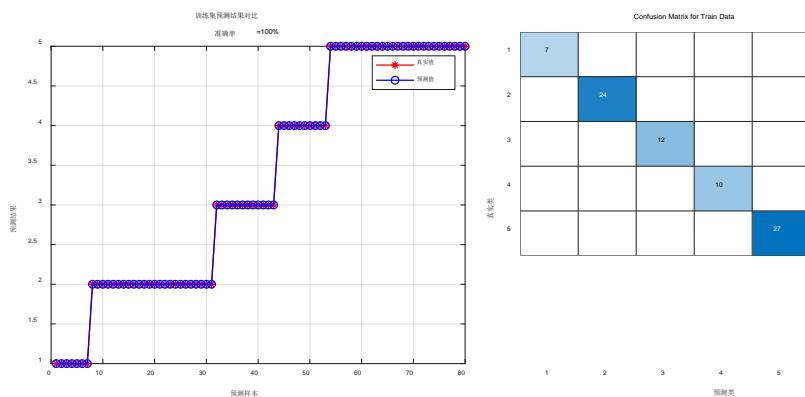


Figure 5. Prediction set accuracy

图 5. 预测集准确率

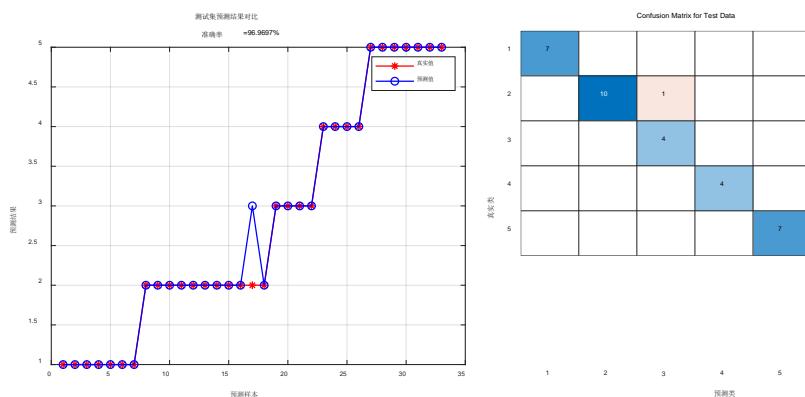


Figure 6. Test set accuracy

图 6. 测试集准确率

5. 结论

本研究通过对电动汽车云端用户数据的分析，深入探讨了电驱动系统中轴、齿轮、轴承等关键部件的失效机理与寿命模型，结合用户驾驶行为，评估了不同用户群体中各个部件的损伤情况与失效风险。研究的主要结论包括：

1) 失效机理与寿命模型的构建：基于对电驱动系统中轴、齿轮、轴承部件失效机理的分析，本研究建立了相应的寿命模型。通过这些模型，我们能够准确预测各部件的损伤程度，并为后续的维护与风险管理提供依据。

2) 风险等级评估与分类：本研究提出了基于不同等级用户的失效风险评估模型。通过对同一等级用户内不同部件的失效风险进行评估，发现即使在同一用户类别内，不同部件的风险等级也存在差异。此外，针对不同等级用户同一部件失效风险等级评估，有助于精确认别高风险用户，进一步提升风险管理的精度。

3) 风险用户等级的定义与应用：研究定义了“风险用户”这一概念，并基于风险用户等级训练了随机森林网络模型。该模型能够有效识别潜在的高风险用户，并为电动汽车的风险管理提供量化的支持。通过这一方法，企业能够更有针对性地进行预防性维护与维修，减少意外故障的发生。

通过以上结论可以看出，本文的研究为电动汽车行业提供了新的思路和方法，尤其是在提高电驱动系统可靠性、优化维护策略、降低运营成本等方面具有重要意义。未来的研究可以进一步深入探讨不同驾驶环境、驾驶模式以及电动汽车使用场景对电驱动系统可靠性的影响，推动智能化和个性化维护服务的实现。

参考文献

- [1] Johannesson, P. (2014) Guide to Load Analysis for Durability in Vehicle Engineering. John Wiley & Sons Inc.
- [2] 赵礼辉, 刘斌, 井清, 等. 轻型商用车全寿命周期目标里程与运行工况研究[J]. 汽车工程, 2018, 40(3): 342-348.
- [3] Zhang, J., Wang, Z., Liu, P., Zhang, Z., Li, X. and Qu, C. (2019) Driving Cycles Construction for Electric Vehicles Considering Road Environment: A Case Study in Beijing. *Applied Energy*, **253**, Article ID: 113514. <https://doi.org/10.1016/j.apenergy.2019.113514>
- [4] 全国汽车标准化技术委员会. 电动汽车用驱动电机系统可靠性试验方法[M]. 北京: 中国标准出版社, 2012.
- [5] 赵礼辉, 王震, 冯金芝, 等. 基于用户大数据的电动汽车驱动系统可靠性试验循环工况构建方法[J]. 机械工程学报, 2021, 57(14): 129-140.
- [6] 严婷婷, 侯卫国, 李明峻, 等. 电动汽车驱动电机可靠性模型的建立[J]. 电子产品可靠性与环境试验, 2020, 38(5): 55-57.
- [7] 冷钢. 复合工况下高速滚动轴承的寿命和可靠性预测[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2012.
- [8] 林程, 梁晟, 宫新乐, 等. 面向极限工况的分布式驱动电动汽车动力学集成控制方法[J]. 汽车工程, 2022, 44(9): 1372-1385.
- [9] 杨年炯, 宋英杰, 肖旺. 某型纯电动汽车动力系统匹配及优化研究[J]. 中国工程机械学报, 2023, 21(6): 557-561, 567.
- [10] 王金. 车用电机控制器寿命评估与寿命提升研究[D]: [硕士学位论文]. 重庆: 重庆大学, 2021.
- [11] 王震. 运行数据下电驱动系统剩余寿命动态预测方法研究[D]: [博士学位论文]. 上海: 上海理工大学, 2024.
- [12] 赵礼辉, 郑松林, 冯金芝. 基于低载强化特性的疲劳寿命估计方法[J]. 机械工程学报, 2013, 49(8): 115-122.