

摘要

智能交通的推广和发展，以及计算机和通信技术不断创新和运用，实现了车辆信息的处理、存储、传输和交换等功能，进而推动了车联网项目的发展。我国车联网项目发展空间广阔，投资规模相对较大，对车联网项目进行风险管理是非常有必要的。车联网项目是一个复杂的系统工程，它涉及到多个领域，在实施过程中，存在诸多问题，车辆故障就是其中一个潜在风险因素。如何对车辆故障进行有效预警和管理是车联网项目亟需解决的问题。通过故障预警降低车辆故障的发生概率，可以减少项目投资风险和损失，并保证项目的顺利实施，提高项目成功的可能性。本文旨在探究基于机器学习的故障预警方法在车联网项目风险管理中的应用，通过分析和挖掘有用数据，提取故障的特征，选择合适的机器学习算法并建立相应的预测模型，实现车联网项目风险管理的目标。

本文研究的内容主要有以下几部分：

(1) 从项目风险管理理论、故障预警研究现状、车联网项目风险管理研究以及机器学习模型在风险管理中的应用等相关理论和研究出发，结合车联网项目的实际开发经验，选取了 70 多个特征用于构建车辆故障预测模型，并使用卡方检验的方法验证了这些特征对建立预测模型的显著相关性。实验结果进一步表明，这些特征的选取对于车辆故障预警研究和实践都具有一定的借鉴意义。

(2) 使用多种机器学习算法构建车辆故障预测模型，包括堆叠集成学习、决策树、支持向量机、逻辑回归、极度随机树、随机森林、自适应提升和梯度提升决策树等。同时，将预测时间提前 1 分钟至 10 分钟，并对这些模型使用多种评价指标进行性能评估。实验结果表明，集成模型的预测效果优于其他单一模型，特别是堆叠集成学习 Stacking 模型的表现最佳。此外，针对不

同的时间点进行预测，结果显示，预测离故障发生时间越近，模型预测的准确率越高。

(3) 进一步验证本文所构建的故障预测模型在实践中的成本收益，使用我国某知名车企车联网项目的最新生产数据进行实证分析。实证结果显示，本文所构建的故障预测模型具有良好的预测性能和准确性，能够对车辆故障进行有效预测和预警。此外，通过对预测模型的应用进行成本收益分析，验证了基于机器学习的故障预警方法在车联网项目故障风险管控中的优势和可行性，为决策者提供了有力的技术支持和信息参考，有助于提高车企故障风险管控的效率和准确性。

根据以上研究内容，本文主要贡献有三点：

(1) 从故障风险管理的角度对车联网项目风险管理进行深入研究。在车联网项目的快速发展过程中，尽管在项目风险管理方面已经有了一些研究成果，但是在故障风险管理方面的研究仍相对匮乏。本文研究实现车辆发动机水温故障预警，可以帮助车主及时发现发动机水温异常，采取相应的措施进行处理，从而有效地降低车联网项目的故障风险。

(2) 采用机器学习方法来进行车联网项目风险管理。虽然机器学习方法已经在风险管理的各个领域得到了广泛应用，但是在车联网项目风险管理方面，其应用还比较有限。本文研究采用机器学习方法建立适用于商用柴油车的发动机车辆故障预测模型，为车联网项目风险管理提供了新的思路和方法。

(3) 为后续车联网项目相关研究提供建模方面的参考和借鉴。本文提出的车辆故障预警方法采用机器学习算法，通过分析和挖掘车辆运行数据的特征，构建车辆故障预测模型，有效预测车辆故障的发生。这一创新性的思路和方法不仅说明了机器学习的适用性，也对机器学习模型的选择和特征的选取提供了新的价值贡献。

关键词：机器学习；故障预警；车联网项目；风险管理

Abstract

The promotion and development of intelligent transportation, as well as continuous innovation and application of computer and communication technologies, has enabled the processing, storage, transmission, and exchange of vehicle information, thereby driving the development of the Internet of Vehicles (IoV) projects. In China, the development space of the IoV projects is broad and the investment scale is relatively large. Therefore, it is necessary to carry out risk management for the IoV projects. The IoV projects is a complex system engineering that involves multiple areas, and there are many problems in the implementation process. Vehicle failures are one potential risk factor. Effective warning and management of vehicle faults is a pressing issue that needs to be addressed in the IoV projects. By reducing the probability of vehicle failures through fault early warning, the investment risk and losses of the project can be reduced, and the smooth implementation and success of the project can be ensured. This article aims to explore the application of machine learning-based fault early warning methods in risk management of IoV projects, and to achieve the goal of risk management in IoV projects by analyzing and mining useful data, extracting the features of faults, selecting appropriate machine learning algorithms, and establishing corresponding prediction models.

The main contents of this study are as follows:

(1) Starting from the project risk management theory, current research on fault early warning, research on risk management of IoV projects, and the application of machine learning models in risk management, this study selects more than 70 features to construct a vehicle failure prediction model and uses the chi-square test method to verify the significant correlation between these features and prediction model construction. The experimental results further indicate that the selection of these features has certain reference significance for the research and practice of vehicle failure warning.

(2) Using various machine learning algorithms to construct vehicle failure prediction models, including stacked ensemble learning, decision trees, support vector machines, logistic regression, extremely randomized trees, random forests, adaptive boosting, and gradient boosting decision trees. Meanwhile, the prediction time is advanced from 1 minute to 10 minutes, and multiple evaluation indicators are used to evaluate the performance of these models. The experimental results show that the integrated model has better prediction performance than other single models, especially the stacking model shows the best performance. In addition, the results show that the closer the predicted time is to the occurrence of the failure, the higher the accuracy of the model prediction.

(3) Further verifying the cost-benefit of the fault prediction model constructed in this study in practice, the latest production data of a well-known Chinese auto enterprise's IoV projects is used for empirical analysis. The empirical results show that the fault prediction model constructed in this study has good prediction performance and accuracy, and can effectively predict and warn vehicle failures. In addition, through cost-benefit analysis of the applied prediction model, the advantages and feasibility of the machine learning-based fault early warning methods in fault risk control of IoV projects are verified, which provides decision-makers with strong technical support and information reference and helps improve the efficiency and accuracy of fault risk control in auto enterprises.

Based on the above research content, this article mainly contributes in three aspects:

(1) This article conducts in-depth research on risk management in the IoV projects from the perspective of fault risk management. Despite some research results in project risk management during the rapid development of the IoV projects, research on fault risk management is still relatively scarce. This study implements a vehicle engine water temperature fault early warning to help car owners detect engine water temperature anomalies in a timely manner and take corresponding measures to reduce the fault risk of the IoV projects effectively.

(2) This article adopts machine learning methods for risk management in the IoV projects. Although machine learning methods have been widely applied in various fields of risk management, their application in the risk management of the IoV projects is still relatively limited. This study uses machine learning methods to

establish a vehicle fault prediction model applicable to commercial diesel engines, providing new ideas and methods for risk management in the IoV projects.

(3) This article provides modeling references for subsequent research related to the IoV projects. The vehicle fault early warning methods proposed in this article uses machine learning algorithms to construct a vehicle fault prediction model by analyzing and mining the characteristics of vehicle operation data, effectively predicting the occurrence of vehicle failures. This innovative idea and method not only demonstrate the applicability of machine learning but also make new valuable contributions to the selection of machine learning models and the selection of features.

Keywords: Machine Learning; Fault Early Warning; Internet of Vehicles Projects; Risk Management

目 录

1 绪论	1
1.1 研究背景和意义	1
1.1.1 研究背景	1
1.1.2 研究意义	4
1.2 研究内容和方法	6
1.2.1 研究内容	6
1.2.2 研究方法	7
1.3 研究结构及思路	8
1.3.1 研究结构	8
1.3.2 研究思路	9
1.4 创新点与主要贡献	10
2 理论基础与文献综述	12
2.1 项目风险管理理论	12
2.1.1 项目风险管理定义	12
2.1.2 项目风险管理流程	13
2.1.3 项目风险管理方法	15
2.2 故障预警研究现状	19
2.3 车联网项目风险管理研究	20
2.4 机器学习在风险管理中的应用研究	22
2.5 文献评述	24
3 数据来源和指标选取	26
3.1 数据来源说明与样本选取	26

3.1.1 数据来源说明	26
3.1.2 样本数据选取	27
3.1.3 样本数据划分	28
3.2 预测目标与特征选取	29
3.2.1 研究指标选取原则	29
3.2.2 车辆故障预测目标选取	29
3.2.3 特征选取	31
3.2.4 研究指标卡方检验	36
3.3 数据预处理	39
3.3.1 缺失值填充	39
3.3.2 标准化处理	39
4 故障预测模型构建	41
4.1 机器学习模型的选取	41
4.2 模型的评价指标	42
4.3 故障预测模型设计与构建	43
4.4 实验结果分析	45
5 成本收益分析	54
5.1 生产数据进一步验证模型性能	54
5.1.1 数据收集说明	54
5.1.2 测试结果分析	55
5.2 故障预测模型应用收益分析	57
5.2.1 评估模型与评价指标选择	57
5.2.2 传统故障风险分析方法对比结果	58
5.2.3 故障预测模型应用收益分析	60
5.3 故障预警模块开发建议	63
5.4 车辆故障风险应对措施	64
6 结论与展望	66
6.1 重要结论	66

6.2 提升车联网项目故障预警管理的建议	67
6.3 局限性与展望	68
参考文献.....	70
致 谢.....	76

1 绪论

本文旨在探究基于机器学习的故障预警方法在车联网项目风险管理中的应用，主要关注的是车联网项目中的车辆故障风险管理，并基于车辆大数据分类采集和实践经验，分析影响车辆故障发生的重要特征指标。然后从实际数据分析的角度出发，引入机器学习算法，对车辆故障进行预测和预警，以更好地进行风险管理和提供支持与指导。本章首先介绍研究背景和意义，接着详细介绍研究内容和方法，之后对研究结构和思路进行简要介绍，最后梳理本文的创新点与主要贡献。

1.1 研究背景和意义

1.1.1 研究背景

随着 5G 技术（第五代通信技术）的不断普及与应用，移动互联和人工智能领域也在快速发展。在此大背景下，智能网联汽车作为“互联网+智能汽车”的重要领域备受瞩目。按照智能网联汽车技术逻辑结构划分，“信息感知”和“决策控制”是其技术核心。汽车通过智能的车机系统搭配先进的车载传感器，在现有移动网络通信技术的支持下，将实现车与万物之间的信息交换和共享。车联网作为智慧交通落地的基石，为智慧交通的推进提供了强有力的支撑。车联网项目作为智能网联汽车领域的重要项目，促使其成为汽车产业智能化发展的重要驱动力之一。

我国高度重视车联网项目的发展，并推动相关政策和战略规划的制定和实施，如表 1-1 所示。发达国家在车联网布局的时间相对于发展中国家提前不少。这些国家通过制定相关标准和法规来完善市场发展机制，同时也通

过国家智慧城市发展战略和智慧交通部署来推动车联网项目的快速落地。以下表 1-2 为例，是美国车联网主要市场的政策和项目。

表 1-1 我国车联网项目相关政策和法规（部分）

发布时间	名称	主要内容
2016 年	《汽车产业中长期发展规划》	智能网联汽车被提升至国家战略高度，要求推进与国际市场同步发展。同时，加速构建法律标准体系以规范行业发展，保障消费者权益。
2017 年	《国家车联网产业标准体系建设指南(智能网联汽车)》	指导思想、基本原则和建设目标被明确，有利于智能网联汽车标准体系建设。
2018 年	《国家车联网产业标准体系建设指南》系列文件	对车联网产业进行了顶层设计，推动了车联网产业技术研发和标准制定的全面发展。
2019 年	《交通强国建设纲要》	推动智能网联汽车研发，旨在形成自主可控的产业链。
2020 年	《智能汽车创新发展战略》	要求建立完整的智能汽车法规标准体系。
2021 年	《“十四五”信息通信行业发展规划》	高速公路和城市道路是蜂窝车联网（C-V2X）的重点覆盖领域。
2022 年	《车联网网络安全和数据安全标准体系建设指南》	到 2025 年，建立完善的车联网安全标准体系，保障数据安全。

表 1-2 美国车联网项目相关政策和法规（部分）

发布时间	名称	主要内容
1991 年	《综合地面交通效率法案》	开始推动智能交通系统（Intelligent Transportation Systems, ITS）的研发和规划。
1999 年	DSRC(专用短程通信技术)	DSRC（专用短程通信技术）频段 5.850-5.925 GHz 被分配并应用。
2004 年	VII(车路协同系统)项目	该项目深化智能交通研究和技术普及，并产生 IntellDrive（智能驾驶）和 CVR（智能互联汽车研究）两个子项目。
2010 年	“ITS-5 年战略 1.0”	详细规划并制定车联网发展目标，助力车联网发展的基础建设。
2015 年	“ITS-5 年战略 2.0”	重点发展方向包括自动驾驶、企业大数据和车路协同等。
2020 年	《智能交通系统(ITS)战略规划 2020-2025》	重点关注 5G 通信技术的推广和应用。
2021 年	《自动驾驶对公路基础设施的影响》	详细分析了自动驾驶的发展和普及，探讨了它对公路基础设施、交通管理设备、商业运输和后台运营系统、多式联运基础设施的影响和意义。

可见,当前全球车联网产业生态正不断完善,汽车联网化渗透日益加强。据全球研究咨询机构埃信华迈(HIS Markit)发布的《中国智能网联市场全球市场发展趋势报告》,到2025年全球市场智能网联汽车渗透率预计将达到60%左右。

在智能交通的推动下,中国车联网用户规模也将逐年扩大,行业渗透率进入快速增长阶段。根据中国信通院发布的《中国车联网白皮书》,到2025年,中国智能汽车市场规模有望突破9600亿元,占全球市场的56.5%,年均复合增速高达36.9%。在此背景下,车联网项目成为车企发展的重要领域之一,对于汽车企业而言,如何在车联网项目中有效管理风险,确保项目顺利实施和运营,成为了亟需解决的问题(商车,2018;解放,2021;李跬,2021)。

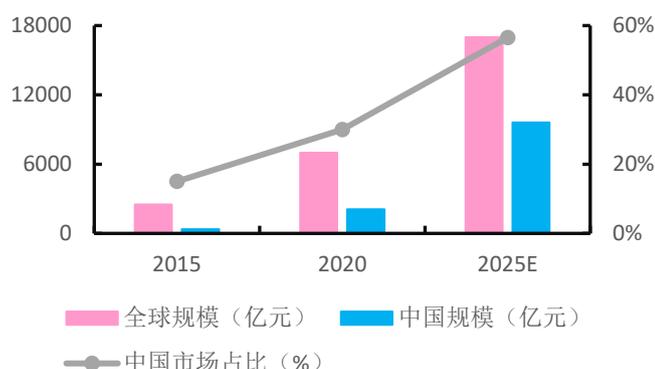


图 1-1 车联网市场规模

注:资料来源《中国车联网白皮书》

随着国内外车联网市场需求的快速增长,许多汽车制造商、互联网企业和软件开发公司都已经投入车联网项目的建设(王春勇等,2018;刘志刚,2018;罗兰贝格管理咨询,2021)。著名的新能源汽车品牌,如特斯拉、小鹏汽车、蔚来汽车等,依靠各自掌握的核心技术和在整个车联网技术中的优势地位,不断加大车联网项目的研发和创新能力;此外,百度、华为、小米、谷歌、苹果等互联网科技领军企业也在加速布局车联网项目,为汽车行业和车联网技术的发展注入强劲动力,同时也加剧各企业在车联网领域的竞争,这对车联网项目的开发和管理提出巨大的挑战(Garg等,2021;Wang等,2021;Gao等,2022)。在这种同行竞争力下,如何有效地管理车联网项目中的风险成为各大车企亟需解决的问题之一,特别是在车辆故障风险管理方

面的问题尤为迫切。道路交通事故中，大约有 2.2% 的事故是由机械故障导致的；机械故障的原因可能是由于驾驶员未按规定对车辆进行例行保养和出车前的检查，使得驾驶员在上路前未能发现车辆存在问题，或者因为车辆本身的机械质量和轮胎气压出现突发性故障（例如制动失效、爆胎等）而导致事故¹。然而，即使驾驶员按规定进行了保养和检查，仍然有可能出现机械故障。因此，为了避免机械故障导致的道路事故发生，驾驶员需要收到及时的故障预警信息，以便及时发现车辆存在的问题并进行维修。因此，故障预警的重要性在于可以帮助驾驶员及时发现车辆存在的问题，避免因机械故障导致的事故发生。

传统模式是基于经验的故障预警方法存在着数据收集、特征提取和模型构建等方面的不足，而基于机器学习的故障预测模型能够更加全面、精准地进行故障预警，提高车辆故障预警的准确率和及时性（Kezia 等，2022）。然而，在实际应用中，基于机器学习的故障预测模型也面临着风险管理上的技术难题。首先，模型的数据采集和处理需要高效且可靠的技术支持，以确保预测模型的准确性和可靠性。其次，模型的性能稳定性需要保障，以应对复杂的道路和交通状况的变化。此外，还需要针对机器学习模型中的黑盒问题进行解释和解决，以提高模型的可理解性和解释性。因此，有效的风险管理策略和措施对于基于机器学习的车辆故障预测模型的实现和应用至关重要。各大车企需要积极研究和探索适合自身需求的风险管理方案，保障基于机器学习的故障预测模型的高效、可靠、稳定、可解释等特性，并将其应用于车联网项目风险管理的实践中。

1.1.2 研究意义

中国城市化进程和经济快速发展，机动车保有量快速增加，商用车作为物流和运输行业的重要载体，也得到广泛的应用（何遥，2017；刘鑫鹏，2018）。2020 年，受国家第六阶段机动车污染物排放标准的升级等政策影响，企业更倾向于选择柴油发动机的商用车，进一步推动商用车行业车联网项目的应用

¹ 参见《道路交通事故损伤构成因素有哪些》，法师兄，2022 年 12 月 17 日，<https://1vlin.baidu.com/question/882662120639906492.html>

和发展。然而，商用车在使用过程中经常会出现各种故障，导致安全、效率和稳定性等方面的问题，已成为制约商用车行业车联网项目发展的瓶颈。因此，商用车行业对车辆故障监测、故障预警等技术有迫切的需求，以降低故障风险和提高车辆的稳定性。对于商用车行业而言，在车联网项目中，研究故障预警系统的优化和应用，具有重要的理论和实践意义。

商用柴油车的发动机是车辆运行的重要构件，其可靠性和健康状况对整个车联网项目的安全和可持续运营至关重要（杜明轩，2019）。本文针对商用柴油车的发动机故障预警进行研究，采用理论研究与实验相结合的方法，分析商用柴油车发动机故障情况，建立最佳的机器学习模型。通过对采集的车辆运行数据进行分析 and 预处理，并使用多种机器学习算法进行训练和优化，最终得到适合商用柴油车的发动机故障预测模型。该模型能够准确预测车辆发动机故障的可能性，为车联网项目中的风险管理提供可靠的预警方法。选题具有理论意义和实践意义如下。

（1）理论意义

首先，从理论意义上讲，本研究为商用柴油车发动机故障预测提供了一种新的理论方法。本研究基于机器学习方法，通过对采集的车辆运行数据进行分析 and 预处理，并使用多种机器学习算法进行训练和优化，建立适合商用柴油车的发动机故障预测模型。

同时，本研究提出的故障特征选取从理论和实际维度出发，充分利用了车辆数据所包含的信息，为车联网项目车辆故障风险管理研究提供了特征选取的理论指导。所选取特征包括发动机空气系统、车身系统、后处理系统、扭矩输出系统、燃油系统、发动机传感器系统、发动机控制系统、变速箱系统、故障诊断系统共 9 个系统维度构成的健康状态基本指标。

此外，从理论上讲，机器学习算法作为一种智能化技术，具有广泛的应用前景和实际意义，因此机器学习算法的研究和应用，可以更好地推动相关领域的智能化发展，这些领域包括智能交通、智慧城市、智能物流等。本文所提出的发动机故障预警方法，通过分析和挖掘车辆运行数据的特征，结合先进的机器学习算法，可以有效地预测车辆故障的发生，为智能交通、智慧城市、智能物流这些相关领域的创新和发展提供了深入思考的理论基础和实践方法。

(2) 实践意义

从实践意义来看，有以下几点：

第一，降低车辆故障率和维修成本

发动机故障预警能够有效减少车辆故障的发生，降低车联网项目中因发动机故障而造成的经济损失和影响，为保证车联网项目平稳不间断运行提供支持。同时，通过对车辆运行状态进行监测和分析，可以优化车辆保养维修策略，降低车辆维修成本。

第二，提高车联网项目的满意度和完成率

车联网项目的成功与否与人们对其的满意度和完成率密切相关。通过发动机故障预警，可以减少车辆故障的发生，提高人们的出行满意度，同时降低因车辆故障导致车联网项目无法达到预期目标的风险，提高项目的完成率。

第三，推动车联网项目的健康发展

车联网的发展目标是实现无人驾驶，并成为物联网系统的中枢平台，对于未来社会的产业形态具有重要影响。预测和管控故障风险可以减少车辆故障的发生，提高人们对于车联网项目的满意度，从而推动车联网项目的健康发展。这对于推动数字化和智能化的发展，提高国家在这一领域的核心竞争力具有重要意义。

1.2 研究内容和方法

1.2.1 研究内容

本文通过将车联网项目、车辆故障风险管理和机器学习模型相结合，实现对车辆故障的精准预测和管理。同时，对机器学习模型应用的成本收益进行分析，证明其在车联网项目故障风险管控中的重要作用。

本文的研究内容主要有以下 3 个方面：

(1) 故障预测模型特征选取

本文首先从项目风险管理理论、故障预警研究现状、车联网项目风险管理研究、以及机器学习在风险管理中的应用研究等理论基础和文献综述展开，然后结合车联网项目的实际开发经验，选取 70 多个特征，并使用卡方检验方

法验证这些特征对建立故障预测模型显著相关，最终选取车辆故障预测模型构建的特征。

（2）故障预测模型构建

在车联网项目中，采集车辆历史运行数据，使用机器学习算法建立分类预测模型，通过使用该模型进行故障风险评估，得出车辆是否会发生故障的结论，从而实现对车辆故障的预测。

（3）模型效果评估与成本收益分析

本文通过准确率(Accuracy)、AUC(Area Under Curve)、精准率(Precision)、召回率(Recall)和 F1_score 等指标对模型的实验效果进行评估，并进行对比分析，选取预测效果更好的模型用于车辆故障风险预测。然后基于我国某知名车企所开展的车联网实际项目对本文构建的故障预测模型进行成本收益分析，验证其在车联网项目故障风险管控中的优势和可行性，为决策者提供更全面的信息。

1.2.2 研究方法

本文主要采用机器学习和文献研究法，旨在深入探究车联网项目中车辆故障风险管理的实践问题，下面对这些研究方法进行详细的分类和总结：

（1）机器学习

在故障预测模型构建、实验结果分析、以及成本收益分析阶段，本文采用机器学习方法（薛祎涵，2018；邢文婷等，2022）。通过采集车联网项目中车辆历史运行数据，并运用机器学习算法建立模型进行故障预测。本文采用包括决策树、SVM、逻辑回归、ExtraTrees、随机森林、XGBoost、LGBM、GBDT、Adaboost、Stacking 这 10 种机器学习模型，并通过对实验进行处理与控制，进行实验并获取结果数据。在成本收益分析阶段，基于我国某知名车企所开展的车联网项目对本文构建的故障预测模型进行效果评估，分析本文研究成果给企业带来的经济效益，最终给出项目风险应对措施，这有利于在实践中验证模型的有效性，并对模型进行不断优化和改进。

（2）文献研究法

在绪论、理论基础与文献综述阶段，本文采用文献研究法。通过对项目风险管理理论进行总结归纳，以及对故障预警研究现状、车联网项目风险管理研究、机器学习在风险管理中的应用研究进行文献梳理，从而得出本文的研究思路，并为本文的研究提供理论基础和参考依据。

1.3 研究结构及思路

1.3.1 研究结构

本文主要是由六个章节构成：

第一章，绪论。主要介绍文章的选题背景、选题意义、研究内容和方法。本文选题的背景是车联网项目的快速发展，但同时也伴随着一系列的故障风险问题。因此，研究如何应用机器学习方法来对故障风险预警，对于车联网项目中的故障风险管理具有重大研究价值。本文使用的研究方法是机器学习，研究思路是构建故障预测模型，通过数据分析和建模来实现故障预警的目标。

第二章，理论基础与文献综述。主要对文章的理论进行梳理，回顾各研究的现状及发展。在本章中，我们将主要介绍以下几个方面的内容：项目风险管理理论、当前故障预警研究现状、车联网项目风险管理研究进展，以及机器学习在风险管理中的应用研究。这些理论和研究都是当前亟需深入探索的领域。通过阐述他们之间的关系以及相互作用，得到本文的研究内容：基于机器学习的故障预警方法在车联网项目风险管理中的应用研究，进一步说明本文的研究意义。

第三章，数据来源和指标选取。主要介绍如何获取用于构建预测模型的车辆历史运行数据，并结合车联网项目的实际开发经验，选取 70 多个特征，然后使用卡方检验方法对这些特征与模型构建的相关性进行验证，说明指标选取的有效性，从而构成本文车辆故障预测模型的指标。这一步是建立故障预测模型的第一步，为后面的模型构建奠定了基础。

第四章，故障预测模型构建。主要使用机器学习算法构建车辆故障预测模型，并对模型进行实验分析。然后对模型的预测准确性、精确度、召回率等指标进行评价，证明机器学习模型可以有效预测车辆故障风险，并且选取

最优模型可以更准确地预测车辆故障风险发生概率，从而能更有效地管理车联网项目故障风险。

第五章，成本收益分析。主要对基于机器学习的故障预测模型进行成本收益分析，总结该模型在实际项目中的应用效果，并探讨其对项目风险管理的实际指导作用。同时，还对模型应用的成本和效益进行详细的分析和评估，以帮助项目管理者更好地理解 and 评估该模型的实际应用效果。

第六章，结论与展望。首先是总结本文的工作，并基于本文的研究结论提出提升车联网项目故障预警管理的建议，然后说明本文研究的局限性，最后展望未来的研究方向，以期对车联网项目中的故障风险管理提供科学严谨的方法和思路。

1.3.2 研究思路

本文的研究思路如下框架图 1-2 所示：

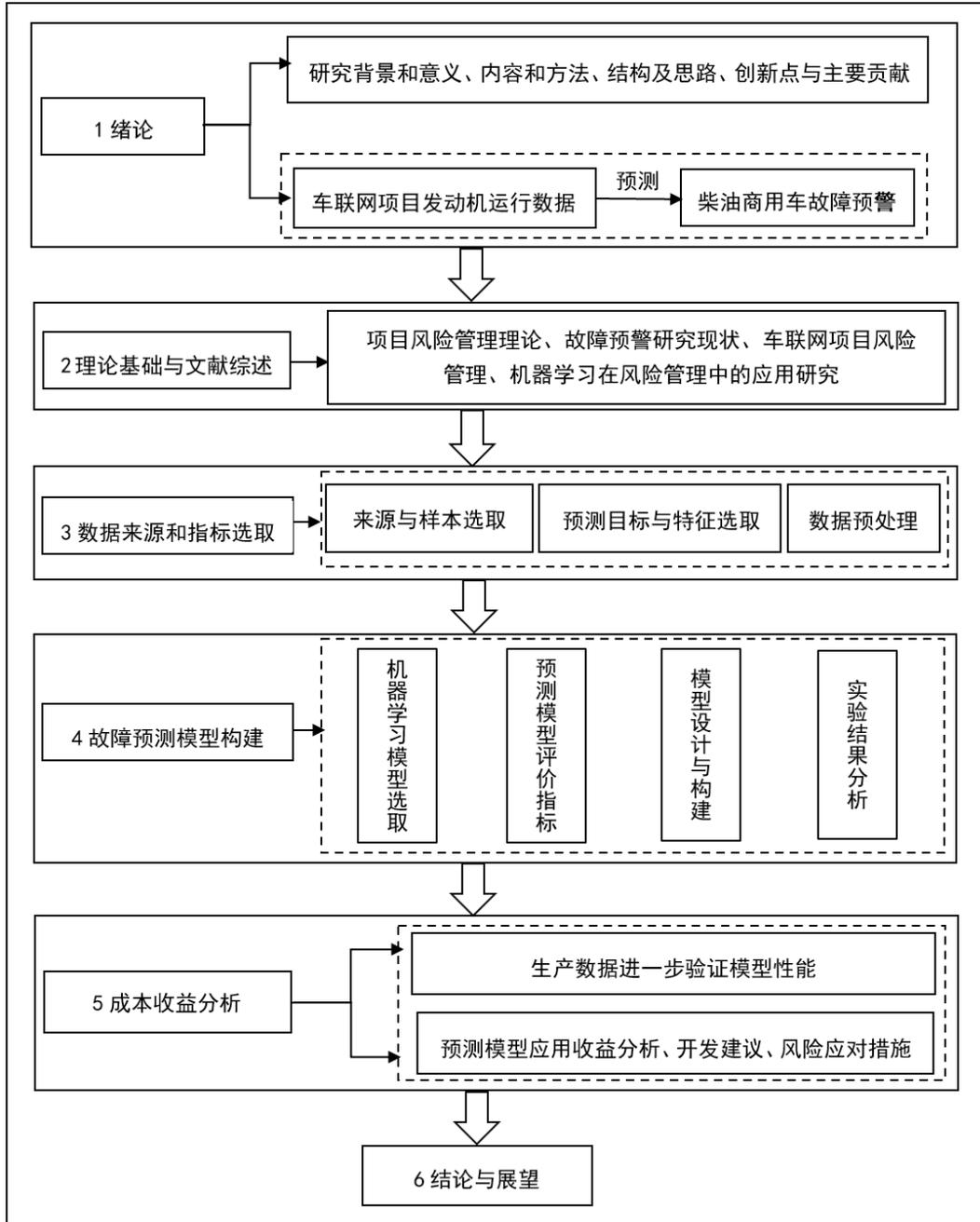


图 1-2 研究框架

1.4 创新点与主要贡献

本文的创新点和主要贡献有三点：

第一，本文从故障风险管理的角度对车联网项目风险管理进行深入研究。在车联网项目的快速发展过程中，尽管在项目风险管理方面已经有了一些研究成果，但是在故障风险管理方面的研究仍相对匮乏。本文研究实现车辆发动机水温故障预警，可以帮助车主及时发现发动机水温异常，采取相应的措施进行处理，从而有效地降低车联网项目的故障风险。

第二，本文研究采用机器学习方法来进行车联网项目风险管理。虽然机器学习方法已经在风险管理的各个领域得到了广泛应用，但是在车联网项目风险管理方面，其应用还比较有限。本文研究采用机器学习方法建立适用于商用柴油车的发动机车辆故障预测模型，为车联网项目风险管理提供了新的思路和方法。

第三，本文构建车辆故障预测模型可以为后续车联网项目相关研究提供建模方面的参考和借鉴。本文提出的车辆故障预警方法采用机器学习算法，通过分析和挖掘车辆运行数据的特征，有效预测车辆故障的发生。这一创新性的思路和方法不仅说明了机器学习的适用性，也对机器学习模型的选择和特征的选取提供了新的价值贡献。

2 理论基础与文献综述

本章首先从项目风险管理的理论出发，对风险管理的定义、流程和方法进行详细论述，为后续车联网项目风险管理研究提供了理论基础。然后，结合当前车联网项目定义和常用风险管理研究方法，以及机器学习运用到项目风险管理和风险管理领域的研究进展进行了归纳和总结，旨在为本文的研究方向提供相关背景和支撑。在理论支撑和应用研究的基础上，本文将基于机器学习的故障预警方法在车联网项目风险管理中的应用作为研究方向，进一步探讨如何应用机器学习算法来提高车联网项目风险管理的效率和准确性。

2.1 项目风险管理理论

2.1.1 项目风险管理定义

风险是指一种不确定的事件或条件，可能对活动、事件或项目产生正面或负面的影响。对于从事活动或项目的人而言，预期结果的不确定性就意味着存在风险（谭志彬等，2017）。

项目风险管理需要考虑如何通过改变以往的行为方式或技术路线来优化未来的活动或项目。在项目风险管理中，风险分为单个项目风险和整体项目风险。单个项目风险会对一个或多个项目目标产生正面或负面的影响。而整体项目风险则是对整个项目的不确定性影响，可能导致项目结果的正面和负面变异区间。无论是单个项目风险还是整体项目风险，风险的系数和发生的时间和地点都是不确定的。这是因为人对客观世界的认识受到各种条件的限制，无法准确预测风险的发生。在不同的开发阶段，项目会随着目标的进展而变化，因此需要做出决策，尽可能防范风险，并将风险的不利性降至最低。

本文研究的是车联网项目中的单个项目风险，即车辆故障风险。如果不能有效地降低车辆故障的发生概率，那么就无法实现项目立项之初与相关方所签订的项目目标，这将导致车联网项目的失败。

以往，大多数项目只关注可能发生或不发生的未来事件的风险，例如客户在设计开发完成后对需求的变更、项目优秀人才的流动、技术研发成本的超支等。但是，现在人们意识到，非事件类风险的识别和管理也非常重要。非事件类风险指的是已规划事件、活动或决策的某些关键方面存在不确定性所导致的变异性风险；例如，生产率可能高于或低于目标值，测试发现的错误数量可能多于或少于预期，或施工阶段可能出现反常的天气情况（美国项目管理协会，2018）。

在车联网项目中，车辆故障预警准确率低于目标值，或者车辆故障发生率高于目标值，都属于车联网项目的非事件类风险。因此，本文所研究的车辆故障预测模型旨在提高车联网项目故障预警的准确率，降低车辆故障的发生率，以期对车联网项目的风险进行有效的管理。

因此，在项目风险管理中需要关注未来事件的风险，也需要注意非事件类风险，才能有效地管理风险。本文所研究的车辆故障预测模型可以帮助我们实现这一目标。

2.1.2 项目风险管理流程

项目风险管理的目标在于提高正面风险的概率和（或）影响，降低负面风险的概率和（或）影响，从而提高项目成功的可能性（美国项目管理协会，2018）。

风险管理各过程包括规划风险管理、识别风险、开展风险分析、规划风险应对、实施风险应对以及监督风险，见图 2-1 所示，每个阶段适用不同的风险管理方法（美国项目管理协会，2018）。

（1）规划风险管理

为了实施项目风险管理活动，首先需要制定一份指导性文件，称为规划风险管理。该文件的主要作用是确定风险管理的技术水平、逻辑方法和实用性，以减少项目风险对组织和其他相关方的影响。

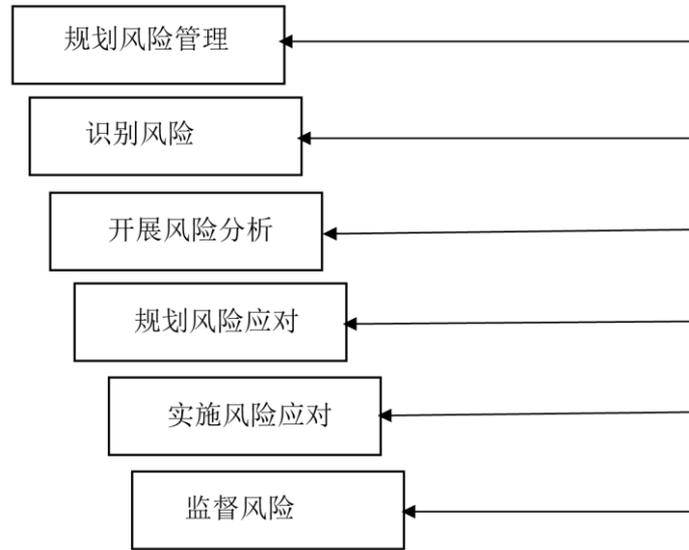


图 2-1 项目风险管理流程

(2) 识别风险

识别风险是在风险管理计划的指导下，识别单个项目和整体项目风险发生的条件，并记录其特征的过程。这个过程还需要收集相关的数据和关键信息，以帮助项目组准确地应对已识别的风险。最终，这些信息将被整合到风险登记册中。

(3) 开展风险分析

进行定性风险分析时，需要对可能出现的风险进行可能性和后果的分析，以便按照风险的影响程度进行排序，并决定哪些风险需要进行进一步的定量分析。该过程的主要目的是关注高风险问题，将其记录在风险登记册中。在某些项目中，可能还需要进行定量风险分析，以量化受风险影响的重要问题。尽可能进行量化分析，以更好地预防风险并确定应对风险的应急决策，避免潜在损失。

(4) 规划风险应对

规划风险应主要是制定应对整体项目风险和单个项目风险的适当方法，对已经识别出来的风险制定应对策略和应对措施，并为相应的商定好的风险决策指定实施责任人。

(5) 实施风险应对

为规避整个项目或单个项目的风险，规划风险管理应当制定适当的应对方法，包括对已识别的风险制定应对策略和应对措施。此外，该过程还应确定相应的风险决策的责任人，以确保及时实施风险应对措施。

（6）监督风险

监控风险是贯穿于整个项目周期的一项工作。它包括识别新的风险，跟踪已经制定的风险应对计划的执行情况，监测剩余风险，并评估已采取的风险应对措施的有效性。在必要时，应提出风险变更请求以及重新评估项目风险。

2.1.3 项目风险管理方法

在项目开发过程中，不同阶段面临的风险因素也不同，因此需要在不同时间段采用不同的风险管理方法。一些常用的项目管理方法包括头脑风暴法、专家调查法、层次分析法、模糊综合评价法以及机器学习方法等。

（1）头脑风暴法

在进行项目管理时，风险管理是非常重要的一环。在项目风险识别阶段，最常用的管理方法之一就是头脑风暴法。这种方法又称为智力激励法或自由思考法，它的初衷是将产生想法的过程与分析、评判想法的过程分离开来。

头脑风暴法已经广泛用在项目风险管理中。韦凯尹（2020）对 W 券商广东分公司财富中心项目进行风险识别时，采用头脑风暴法和 SWOT 分析识别出了运营、市场、监管和系统性等四类风险，共 13 个风险因素。李凡（2021）针对 DW 食品公司开发计量器具管理平台的项目进行风险管理研究时，使用了头脑风暴法，通过成员之间充分交流和沟通，对项目技术、质量、成本和进度等方面进行了全面的风险识别。盛珈琨（2021）针对公司新建项目进行了风险管理规划，在风险识别阶段，使用了头脑风暴法识别了市场、技术、HSE、质量、成本、资金和进度等方面的风险。

头脑风暴法是一种广泛应用于项目管理中的创意产生方法，其适用于风险识别和评估。几乎所有类型的项目都可以使用头脑风暴法进行项目风险管理。然而，头脑风暴法的结果通常是概括性的，而不是详细的或量化的。在头脑风暴会议中，参与者会面对面地交流意见，但他们的意见可能会受到他

人的影响，这可能会影响到头脑风暴法的有效性。因此，在使用头脑风暴法时，需要认识到其局限性，并结合其他工具和技术来实现更全面和精确的风险管理。

（2）专家调查法

专家调查法又叫德尔菲法，是一种专家意见征询方法，由美国兰德公司于 1946 年创始实行。该方法通过匿名函询的方式对专家群体进行反复征求意见，以期望最终得到一致的结论。

目前，德尔菲法已经成为项目风险管理中广泛应用的一种方法。W 银行使用德尔菲法和层次分析法等风险识别工具，评估投资风险并计算风险系数，制定专项方案进行整改来提高企业的整体投资回报率，以保障国有资产增值（李晓璇等，2021）。为识别南市历城区 H 房地产项目的风险因素，采用德尔菲法进行研究，将项目流程分为投资决策、规划设计、工程建设、销售回款和交付使用五个阶段，并识别出了每个阶段的 19 个二级风险因素，这一方法可更全面、有针对性地评估风险、降低风险并提升项目成功率（边凯，2022）。

德尔菲法是一种广泛应用于风险识别、评估和管理的方法，适用于各种类型的项目。但是该方法存在一定的主观性，只能得到一个大致结果，无法提供详细或量化的结果。此外，德尔菲法实施起来非常繁琐，需要多次的调查才能得出最终结果。因此，对于风险管理要求不高或者投入资金较少的项目，可以考虑使用德尔菲法进行风险管理。但对于风险管理要求较高或投入资金较多的项目，可能需要使用更为精细和复杂的风险管理方法。

（3）层次分析法

层次分析法（AHP）是一种决策方法，它通过将决策问题分解成目标、准则、方案等层次，对各层次之间的关系进行定性和定量分析，最终得出综合评价结果。

层次分析法在项目风险管理中被广泛应用，其可帮助项目管理者快速确定关键因素，分析和解决问题，提高风险管理效率。Zheng 等（2012）基于层次分析法建立了包含三个因素和十个子因素的安全评价框架来评估热湿环境下的工作安全，并确定综合安全指数、安全等级和预警等级，具有工程实用性和有效性。Zhong 等（2019）针对定制生产项目，应用层次分析法识别 22 个三级因子，并最终确定了 TOP10 因子，这些关键因素的确定帮助项目经理

快速找到问题的核心，并在项目管理中进行分析 and 解决，提高了项目管理的效率与质量。孙仓龙（2022）针对某改造工程 EPC 总承包项目的风险管理，采用层次分析法进行量化计算和评估，通过准确确定 EPC 总承包项目中风险因素的主要权重，项目管理者可以有针对性地实施风险控制，并为其他类似工程提供经验借鉴。刘阳（2022）对汽车改装企业 ERP 实施项目的风险因素采用层次分析法进行分析，得出结论：安全风险对整体系统风险产生的影响最大，准备风险对整体系统风险产生的影响最小；在方案层面，需要重视的风险因素包括数据转换风险、进度风险、网络安全风险和实施效果评价风险；这一研究结果有助于项目管理者识别风险因素并有针对性地实施风险控制措施，同时也可为类似项目提供有益的借鉴。

层次分析法是一种系统化、方便实用且所需的定量数据较少的决策方法，但在使用时需要注意，由于该方法利用的数据中定性因素成分较大，当判断矩阵的规模较大时，计算会变得困难。

（4）模糊综合评价法

模糊综合评价法是一种将复杂的定性评价转化为定量评价的方法，其依据模糊数学理论进行运算，能够考虑评价指标之间相互影响和不确定性因素，为决策提供全面客观的参考。该方法具有结果清晰、系统性强等优点，适用于解决各种非确定性问题。通过使用隶属度理论对受多种因素制约的事物或对象进行综合评价，可以较好地解决难以量化的问题。

项目管理中常用模糊综合评价法进行决策，因为它能够较好地处理非确定性的因素，帮助管理者在不确定情况下做出决策，提高项目的成功率。Zheng（2013）为了更准确地评估电力工程项目风险，使用了多层次模糊综合评价模型，以某电力建设项目为例进行了实际应用和分析。Gao 等（2013）针对中国建筑市场存在的风险问题，通过系统性的应用模糊综合评判法，从风险因素的选取及其权重、专家评价得分等方面进行评估，并对风险应对方案的可靠性进行了分析；研究结果表明，模糊综合评判法在风险解决方案的选择上具有较高的可行性，通过综合考虑专家经验和定量分析，提高了评价结果的科学性、合理性和准确性。许宗武等（2022）针对印尼物探项目分包商管理过程中的风险因素、风险原因及风险后果，使用了层次分析法和模糊综合

评价法进行客观、科学和全面的评估，以便更好地识别、分析和控制风险，确保项目管理的有效性和成功性。

虽然模糊综合评价法可以将定性指标转化为定量评价，但是在应对大量项目风险因素时，可能会出现权重过小、分辨能力低、评估结果不明确的问题，这被称为“超模糊现象”，可能导致评估失败。

（5）机器学习方法

机器学习是一种科学，它开发了一些算法和统计模型，使计算机系统可以在没有人类指令的情况下执行任务，通过利用大量历史数据，识别出数据模式，从而提高计算机系统预测结果的准确性。例如，机器学习通过处理大量图形数据，提供图像处理和分析服务（Wei, 2021），机器学习也可以通过分析大量健康数据，为医生的诊断和治疗提供支持等。

目前，机器学习算法在项目风险管理领域备受关注，是未来发展的研究方向之一。使用机器学习技术预测建筑项目延误风险，并基于客观数据源进行风险分析和预测，经评估，朴素贝叶斯模型在数据集上表现更佳；通过此项工作，利用机器学习优势来推动基于数据的决策，针对项目风险管理策略的固有风险因素，提高了项目管理的主动性和决策的准确性（Gondia, 2020）。对中国企业在投资中东欧国家的重大项目风险进行分析时，利用机器学习方法，包括线性判别分析、二次判别分析、朴素贝叶斯分类、基于树的方法（如决策树、套袋和随机森林）、多分类逻辑回归、多元逻辑回归、岭回归和套索回归、主成分回归和偏最小二乘等多种模型，该研究展示了各解释变量在预测中的影响和重要性排序，是前人研究的学习、总结和补充，也为未来对海外投资项目的系统性风险预测和管理提供了方法论，具有实践应用价值和借鉴意义（刘一凡，2021）。使用机器学习方法预警高风险项目，提出了一种基于距离评判和支持向量数据描述的智能预警模型，该模型不需要过多甚至不需要高风险项目样本即可实现高风险项目风险预警；经案例分析验证，该方法可以为项目风险管理提供新的思路和方法（王巍，2008）。对于裕农快贷的借款人进行贷款发放的预测评估时，使用逻辑回归、支持向量机、类神经网络以及 K-means 聚类等机器学习模型，最终，提出了一种混合模型，该模型结合了有监督的聚类算法和回归算法，使银行能更加精准地投放贷款，从而降低贷款违约的风险（杨晴，2022）。对信托公司在互联网借贷项目中，

借助机器学习中的常见模型，对借款人违约所受到的各种影响因素进行了综合评估；同时，对这些模型的可信度和稳定性进行了评估，为信托公司今后在开展互联网借贷项目时提供了风险管理机器学习模型和项目改进建议（刘斯玮，2022）。

可见，机器学习技术在企业项目中的应用，可以帮助企业发展、开拓新的收入来源，解决难题，提高决策效率。数据是商业决策的核心驱动力，但传统的数据处理流程通常涉及来自不同来源的数据，如客户反馈、员工和财务数据等。机器学习研究对这一流程进行了自动化和优化，通过使用高速数据分析软件，可以更快地获得项目成果。

在本文研究中，旨在于使用机器学习算法构建车辆故障预测模型，以降低车联网项目中的故障发生风险，确保实现车联网项目的目标。这一模型的构建对于保障车联网系统的安全运行和提高用户满意度都具有重要意义。

2.2 故障预警研究现状

故障诊断技术的研究经历了多个发展阶段。早期的故障诊断起源于 19 世纪末至 20 世纪初期，使用了人工的手段，以经验和直觉为主要依据。由于设备结构简单，这种方法可以较好地发挥作用。

从 20 世纪初期一直到 70 年代，随着仪器设备技术的进步，常规仪器诊断逐渐被引入。这个阶段的故障诊断主要依赖于各种检测仪器和设备，例如振动分析仪、红外线热成像仪、电子万用表等。这些设备的出现使得故障的检测更加准确和快速。

随着 80 年代后期电子电路技术的更新，设备结构更加复杂和精密，传统的人工和常规诊断已无法满足爆炸性增长的数据信息处理需求。机器学习技术在这个时候应运而生。机器学习技术是基于计算机技术、数据处理技术等各种方法相结合的一种新型诊断方法。它利用计算机对海量数据进行快速处理和分析，通过数据挖掘和机器学习算法进行故障预测和诊断（匡旭等，2016；郭少丹，2020）。机器学习技术不仅能够大幅提高诊断的准确度，同时还能够大大缩短诊断时间，节约人力成本。

目前，机器学习算法已开始在故障预警方面得到应用，以传感器数据为基础，通过训练模型实现设备状态实时监测和故障诊断(Rezamand 等, 2020; 贺东旭, 2022; 黄日新等, 2022)。更多机器学习故障预警相关研究见下表 2-1 所示。

表 2-1 机器学习故障预警相关研究

作者	研究对象	研究方法	研究成果
Bangalore 等 (2015)	风力涡轮机齿轮箱轴承	人工神经网络 (ANN)	实现轴承磨损预测，提高维护效率
薛祎涵 (2018)	公交车发动机	改进的随机森林算法	提供工作决策和数据依据
刘建强等 (2019)	地铁车辆轴承	均值加权融合法	提高地铁车厢轴承故障诊断的准确率
马立玲等 (2020)	车辆故障预警	不平衡数据分类支持向量机算法	有效提升故障诊断的效率和准确性
Chen 等 (2020)	风力机俯仰系统	遗传算法优化 BP 神经网络算法	能满足故障预警需求
秦春林 (2021)	铁路列车轴温	改进的支持向量机	准确率高达 97.67%
Zhang 等 (2021)	提升机故障预警	基于大数据	达到故障预测的要求
Wang 等 (2021)	停电故障	支持向量机(SVM)二义分类器	有助于电网系统的规划和维护
Zhao 等 (2021)	故障预测	提高建模数据质量	提高故障预测模型的误判和性能稳定性
刘勇等 (2022)	航空发动机预测	数据降维的核主成分分析(Kernel PCA)原理和机器学习中决策树 (GBDT) 算法	泛化性能强劲，F1 分数高于 0.99，ROC 曲线面积值达到了 0.998

随着计算机技术和数据处理技术的不断发展，机器学习技术将会得到更广泛的应用，它将成为未来故障诊断的主要方法之一。然而，基于机器学习算法的车辆故障风险评估与预警应用研究还比较缺乏，尤其是在车联网项目中，如何利用机器学习算法预警潜在故障的车辆，是一个非常值得结合实际场景进行验证和分析的问题。最终通过机器学习算法构建最优的故障预测模型，是本文研究的目的和意义所在。

2.3 车联网项目风险管理研究

目前，传统的风险管理方法在车联网项目中仍然是常用的方法。这些方

法包括头脑风暴法、专家调查法、层次分析法和模糊综合评价法等。然而，随着机器学习技术的发展，越来越多的研究者开始使用机器学习方法来进行项目风险管理研究，以提高风险管理的效率和准确性（Sharma 等，2021；Siddiqui 等，2021）。但是，在车联网项目风险管理方面，对于使用机器学习方法的研究还相对较少。因此，针对车联网项目的机器学习研究在提高风险管理效率和准确性方面具有重要价值。

基于移动云计算技术的车联网项目中，研究使用风险分解结构来判断风险，并采用模糊综合评价法对风险进行定量分析，最后，使用风险定量分析的结果来制定风险应对策略，为该项目的风险管理提供了理论基础（高雪松，2012）。针对汽车创新产品项目开发中的风险管理，研究将优秀的风险管理理论应用于车联网项目开发中，主要采用风险评估检查表法和工作分解结构法进行风险识别，采用风险影响矩阵法和专家评分法进行定性风险分析，通过风险转移、风险规避、风险减轻和风险自留等应对策略进行风险应对，对遗留风险进行风险监控，在此基础上，提出了适用于该项目的最佳风险管理方案，并进行总结（赵任驰，2016）。为了更好地管理 G 公司的车联网项目风险，采用多种风险管理方法，包括分解分析法、头脑风暴法和风险专家调查法，以确保对车联网项目开发、建设和运行阶段的风险进行全面识别，还使用了网络分析法（Analytic Network Process）来确定项目风险的占比，并结合模糊综合评价法对项目风险进行综合评定，记录了项目风险的综合评价结果，最后，对项目风险进行排序，并为不同等级的风险提供相应的预应对措施和负责人，这些方法的运用，为 G 公司的车联网项目风险管理提供了有效的理论基础和实践指导（高见，2017）。研究提出一种基于改进的 FMEA（失效模式与影响分析）的车联网项目风险管理模式，该模式采用失效模式进行项目风险识别，避免了传统项目风险评估指标（如发生度、严重度、检测难度等）的固定性限制，针对车联网项目的特点，运用模糊综合评价对项目进行风险评估，并提出相应的风险管理策略，该模式为车联网项目风险管理提供了新思路和方法（杜宁，2019）。针对 BA 汽车公司车联网项目进行系统性的风险分析，结合公司原有的整车开发项目流程，全面识别该项目开发过程中可能出现的各类风险，首先，采用头脑风暴法、专家调查法和工作结构分解法等工具，对项目风险进行了统计和清单化，随后，使用风险概率影响

矩阵对风险进行了等级划分,并从专家打分结果中计算具体风险值,接下来,采用层次分析法对风险进行总排序,并最终使用模糊综合评价法对项目的总体风险进行评估(李家骏,2020)。为了研究车辆技术条件对实时驾驶风险管理的影响,研究提出一种基于模糊理论的综合驾驶支持系统,并应用于实时风险管理,该系统通过模拟考虑各个参数对驾驶风险评估的影响,并在主要风险不超过0.2时标记驾驶状态为安全,当出现风险时产生警报并采取适当的行动,此外,当VTC(Vehicle-to-Cloud)不超过0.5时,即使其他参数正常,也会产生告警(Bylykbashi等,2021)。

当前,对于车联网项目风险管理的研究已经相对较为广泛,包括传统的方法(如头脑风暴法、分解分析法、专家调查法、模糊综合评价法等)以及使用机器学习的方法。不过,对于后者的研究仍相对较少,机器学习方法在车联网项目中的研究和应用发展值得进一步探究。

2.4 机器学习在风险管理中的应用研究

机器学习作为一种快速发展的技术,在各个领域都得到了广泛的应用。在风险管理中,机器学习的应用不仅可以提高风险评估的精度,还可以提高高风险预测的准确性,并且可以通过自动化的方式大大提高风险管理的效率(Rimal等,2022)。因此,无论是汽车领域还是其他领域的风险管理,机器学习都可以帮助组织更好地理解 and 应对风险,从而降低潜在的损失和成本。

在工业领域,为了提高我国天然气进口预测的准确性和可靠性,研究采用了多种机器学习方法,并对比它们的拟合精度,从而选择出最优模型进行预测;具体而言,构建了多变量灰色GM(1,N)模型、支持向量机回归(SVR)和卷积神经网络(CNN)三种机器学习模型,针对2006~2020年我国天然气进口数据进行了拟合,实验结果表明,机器学习方法能够有效降低天然气进口预测的风险,同时为我国天然气进口及风险管理提供了科学的依据,未来随着数据量和机器学习技术的不断提升,机器学习方法将在天然气进口预测中发挥越来越重要的作用,为我国天然气市场的健康发展提供强有力的支持(邢文婷等,2022)。为即将建设的分布式光伏电站EPC项目建立一套有参考价值的风险评估体系和风险管控策略,研究采用引入模糊隶属度的最小二

乘支持向量机算法构建一套分布式光伏电站 EPC 项目风险评估模型,采用改进的网格搜索与交叉验证的方法对参数进行调优,最终将该模型命名为改进网格搜索的 FLS-SVM 模型;经过实验验证,FLS-SVM 模型比传统的 LS-SVM 模型具有更强的性能,具有很高的应用价值,该研究为分布式光伏电站 EPC 项目的风险评估和管控提供了有效的工具和方法(马帅,2018)。

在金融领域,为了对房地产行业企业所得税税收风险进行识别,基于内蒙古税务局相关数据,通过构建随机森林模型并进行数据模型分析,按照数据挖掘流程对模型的应用进行了分析,包括业务解释、数据采集和预处理、模型构建、模型检验、结果输出以及模型评价,实现了对税收风险的量化衡量和重要性排序;检验结果显示,所建立的随机森林模型能够准确识别出企业所得税申报风险较高的企业(李茹,2022)。为了探究银行风险管理中的反洗钱和恐怖融资预防问题,研究检验有监督(分类、回归)、无监督(聚类、异常检测)以及混合机器学习模型及算法,根据底层理论逻辑和比较结果高度确定了机器学习算法之间的选择(Prisznyák 等,2022)。

在医疗领域,为了预测化妆品的质量和风险,运用机器学习方法对广东药检所历年的化妆品检验数据进行分析,构建多个机器学习风险预警模型,并通过比较它们的预测精度和效果来评估和选择最佳的预测模型算法,最终,选择了最优的机器学习模型作为化妆品风险预警模型算法;这种方法为预测化妆品的质量和风险提供了更加准确、高效和可靠的解决方案(王林,2022)。

在交通领域,为了建立道路货运企业的风险评估模型,研究选择随机森林算法、决策树算法和有序 Logit 算法作为研究对象,并从预测精度、可靠性和可解释性三个方面进行了比较分析;随机森林算法通过构建多棵决策树来进行预测,具有较高的准确率和可靠性,且易于解释和应用,因此在道路货运企业风险评估方面表现出更好的性能;此外,结合道路货运企业市场特点和行业性质,研究进一步发现随机森林算法更适合应用于道路货运企业的风险评估工作中;这一研究结果为道路货运企业提供了一种科学的、可靠的、实用的风险评估方法,有望为道路货运企业的风险管理和决策提供有力支持(左右铭,2022)。为了构建高铁项目社会稳定风险生成网络结构,研究采用多种方法,首先,利用风险生成机理的概念以及相关案例文献中的结论,结合机器学习样本使用问卷的调查表,建立了一个初步的风险网络,接着,

采用专家经验与机器学习相结合的方式，进一步完善了风险网络，最后，通过专家访谈法对局部因果关系进行修正，得出了风险生成过程中风险因素之间、风险事件之间以及风险因素到风险事件的关联关系，从而构建了完整的高铁社会稳定风险生成网络结构；该研究方法具有创新性和实用性，为高铁项目社会稳定风险评估提供了有效的参考（薛雨桐，2021）。

在其他更多领域，针对概率失效风险评估，研究专门设计两种机器学习方法，旨在准确有效地预测应力强度因子；对于第一种方法，进行了三个测试，并对高斯过程回归、树结构模型和人工神经网络的插值和外推能力进行了评估和比较，结果表明，人工神经网络和极度随机树表现更好，这为应力强度因子的准确预测提供了强有力的支持；第二种方法，提出混合模型，并与权重函数进行了比较，实验结果表明，混合模型可以提高 SIF 的计算精度 5% 至 35% 不等；此外，还采用真实的航空发动机轮盘，使用机器学习方法进行轮盘寿命计算，结果显示误差小于 20%，表明了该方法在实际应用中的可行性和准确性；研究所提出的机器学习方法在概率失效风险评估中具有很好的应用前景和发展潜力，可以为相关领域的研究和应用提供重要的参考和支持（Xu 等，2022）。

机器学习在风险管理中的应用越来越受到各行各业的青睐，与传统方法相比，机器学习具备显著的效率和商业价值，可以通过智能算法处理大量数据，实现精准的风险识别和预测，为企业降低成本和增加竞争力提供有效支持。在当前研究中，支持向量机、逻辑回归和随机森林等算法已经被广泛应用于建立风险预测和评估模型，未来还会有更多新算法和模型应用于项目风险管理中。在上一小节 2.3 车联网项目风险管理研究可知，机器学习方法在车联网项目中正成为一个新兴研究方向，并逐渐得到完善和拓展，为车联网项目风险管理提供有力支持，具有巨大的发展潜力和研究价值。

2.5 文献评述

本章通过对项目风险管理理论、故障预警研究现状、车联网项目风险管理、以及机器学习在风险管理中的应用研究进行了梳理和归纳。综合分析相关文献，本文得出以下结论：

首先，车联网项目作为一个快速发展的领域，其风险管理显得尤为重要。虽然在项目风险管理方面已经有了一些研究成果，但是在故障风险管理方面的研究还比较匮乏。因此，未来车联网项目风险管理的研究可以从故障风险管理角度进行深入研究。

其次，机器学习算法作为一种先进的技术手段，已经在风险管理各个领域得到了广泛应用。然而，在车联网项目风险管理方面，机器学习算法的应用还比较有限。因此，在未来的研究中，可以探索如何运用机器学习算法来进行车联网项目风险管理。

最后，综合上述两点，可以将车联网项目、故障风险管理、机器学习三个关键词联系起来，构成本文的研究方向：基于机器学习的故障预警方法在车联网项目风险管理中的应用研究，是一个有价值的探索方向，该方向的研究成果不仅可以提高车联网项目的管理效率和安全性，还可以为其他领域的项目风险管理提供借鉴和参考。

3 数据来源和指标选取

本章主要介绍在车联网项目中如何获取构建故障预测模型的车辆运行数据，并根据发动机编号和运行时间匹配，将发动机故障数据标记为 1，其他正常运行数据标记为 0。在获得数据之后，本文选取 70 多个特征，并使用卡方检验方法验证这些特征与模型构建的相关性，进一步证明所选取指标的有效性。最后，对所选取的样本数据进行缺失值和标准化处理。本章内容是建立模型的关键步骤，为后续的模型构建奠定了坚实的基础。

3.1 数据来源说明与样本选取

3.1.1 数据来源说明

在对车辆故障预警进行研究时，数据来源是一个至关重要的问题。在本研究中，我们得到的车辆数据来源于我国知名车企 A 企业。

A 企业是一家以商用车生产为主的企业，致力于为全球客户提供安全、高效、环保的车辆解决方案。随着社会的不断发展和技术的不断进步，车辆故障的处理和管理成为制约企业发展的重要因素。因此，为了提高车辆的运行效率和降低维修成本，A 企业决定实施车联网项目。

开展该项目的背景是，随着汽车制造技术的发展和普及，车辆中的传感器和电子控制单元越来越多，车辆的智能化程度也不断提高。但是，在实际运行过程中，车辆故障仍然是一个无法避免的问题，不仅影响了车辆的正常运行，还会导致额外的维修成本和停机时间，甚至对人员和环境造成安全隐患。因此，通过车联网技术对车辆进行实时监测和故障预测，成为降低车辆故障率、提高车辆运行效率的有效途径。

在这种背景下，A 企业推行了车联网项目，通过车载传感器采集车辆的运行数据，实时监测和分析车辆的运行状态，实现对车辆故障的预测和预警。该项目从安装传感器、采集数据、数据分析和处理等环节入手，通过协同作用，实现了对车辆故障预测的目标。

使用 A 企业的数据进行故障预警具有重要的行业价值。首先，该企业是商用车行业的领先企业之一，其生产的车辆应用范围广泛，有着较高的市场份额。其次，A 企业在实施车联网项目方面投入了大量人力、物力和财力，具有较高的技术水平和实践经验。因此，通过使用 A 企业的数据，可以对车辆故障预测模型进行充分的验证和优化，进一步提高模型的准确度和可靠性，为商用车行业提供更好的服务和支持。同时，该研究可以为其他车企在实施车联网项目方面提供有价值的参考和借鉴。

3.1.2 样本数据选取

为了适应行业市场竞争需求，更好服务客户，A 企业从 2020 年开始规划并实施车联网项目。通过车载终端，实时获取发动机运转数据，并通过移动网络上传到项目服务器，A 企业结合对发动机专业知识的理解，对车联网数据处理分析，再为用户提供个性化路谱等增值服务。

通过采集 A 企业车联网项目近 2 年的数据，即 2020 年 9 月 1 日到 2022 年 8 月 31 日发动机历史运行数据，以及这段时间发动机故障信息的数据，然后用这 2 个数据集的发动机编号以及运行时间进行匹配，将发动机发生故障的运行数据进行打标签为 1，其他非故障数据打标签为 0，构成了本研究的整个数据集。故障发生时刻的指标数据更具有现实意义（薛祎涵，2018），因此本文选取故障发生前 1 分钟、前 2 分钟、前 3 分钟、前 4 分钟、前 5 分钟、前 6 分钟、前 7 分钟、前 8 分钟、前 9 分钟和前 10 分钟的数据作为 10 个样本数据集，即打标签为 1 的故障数据。最后，用于本文实验研究的一共是 10 个数据集，提前 1 分钟进行故障预测的数据集记为 T-1，同理其他数据集为，T-2、T-3、T-4、T-5、T-6、T-7、T-8、T-9、T-10。这 10 个数据集，每个数据集 9600 条数据，包含故障数据是 2400 条，非故障数据是 7200 条，即故障

数据和非故障数据的比例是 1:3，见表 3-1 所示，这是符合实际意义的，因为在实际中非故障数据的占比才是多数的，更具有学术研究价值。

另外，需要说明的是，为了控制变量，本文选取的数据都是同一品牌的车辆，以及同一电控系统的发动机，选取的故障类型是发动机水温故障。

表 3-1 各数据集的故障车辆与非故障车辆数量和比例

样本集	故障车辆	非故障车辆	比例
T-1	2400	7200	1:3
T-2	2400	7200	1:3
T-3	2400	7200	1:3
T-4	2400	7200	1:3
T-5	2400	7200	1:3
T-6	2400	7200	1:3
T-7	2400	7200	1:3
T-8	2400	7200	1:3
T-9	2400	7200	1:3
T-10	2400	7200	1:3

3.1.3 样本数据划分

在确定完样本数据后，需要对样本数据进行划分，将样本数据集划分成两个互不相交的子集，一个作为训练集用于训练，另一个作为测试集对训练模型的评估，从而对模型进行验证，即在这之前需要划分好训练集与测试集。一般来说是将总数据集的 80% 划分为训练集，20% 划分为测试集。本文每个数据集是 9600 条数据，故在本文研究中每个数据集的训练集是 7680 条数据，测试集数据是 1920 条数据，即训练集和测试集的比例是 4:1。各个数据集训练集和测试集划分数量和比例，如表 3-2 所示。

表 3-2 各数据集的训练集和测试集划分数量和比例

样本集	训练集	测试集	比例
T-1	7680	1920	4:1
T-2	7680	1920	4:1
T-3	7680	1920	4:1
T-4	7680	1920	4:1
T-5	7680	1920	4:1
T-6	7680	1920	4:1
T-7	7680	1920	4:1
T-8	7680	1920	4:1
T-9	7680	1920	4:1
T-10	7680	1920	4:1

3.2 预测目标与特征选取

3.2.1 研究指标选取原则

发动机故障预测模型的指标选择目前没有统一定义，并且相关的研究也不多，所以可尝试选择的指标范围就会变得很大。但是选取越多的指标并不意味着模型的效果就会越好，有些指标之间存在相关性，或者有些指标缺失值非常严重，或者有些指标的意义不大。因此，指标的选取要科学全面，不能盲目追求指标数量的多少。

在训练模型前，选取发动机故障预警指标是提高车联网项目发动机故障预测模型性能的关键条件，根据以往预警类研究中指标的选取原则，再结合所学的机器学习特征选取原则，得出本文指标选取原则：

(1)波动性。特征取值发生变化的情况很小，说明该特征的取值很稳定，可以近似理解成该特征的每个值都接近，这样的特征对模型是没有任何效果，是不具备区分度的。

(2)相关性。特征和目标结果的相关性大小，本文使用卡方检验的方法说明选取的特征与故障预测结果的相关性。

(3)真实性。选取的指标必须是真实的数值，并挑选有说服力的特征字段，这些参数可以经受多次反复试验稳定反馈发动机状态的，不能根据主观理解或者其它附加要求而任意变换。

(4)全面性。车联网项目中影响车辆故障的因素可能涉及面广，在研究此类问题前选择特征指标要结合实际尽可能多的全面考虑进去，不能通过单一的指标来预测车辆是否存在故障。

(5)易获取性。指标必须能够普遍环境下都能获取，通过大量数据样本，更具说服力。

3.2.2 车辆故障预测目标选取

柴油发动机是商用车的大心脏，发动机水温是维持发动机正常运行的重要指标。水温过热或过冷，都会导致汽车发生水温故障。发动机过热会导

致喷油系统无法正常工作，会让发动机出现失去动力、爆缸等问题，严重影响汽车的使用寿命。发动机水温过低，则会导致汽车难以启动或降低发动机的能效，增加汽车油耗。

发动机水温故障是车辆故障中比较常见的故障之一，其发生的频率也较高，占到故障比重的 30% 左右，若水温故障不能得到及时解决，会导致发动机熔断，发动机熔断需要更多的工作量和零部件来修复，发动机熔断可能还会导致更换整个发动机，造成更大的损失。发动机熔断情况严重的还可能对车辆其他部件造成损坏，例如冷却系统、排气系统等，也将增加维修成本，这对车主而言是一个相当昂贵的支出。

因此，水温故障预警的重要性不言而喻。它可以帮助车主及时发现发动机水温异常，防止因发动机过热造成的危险情况发生，提高汽车的行驶安全性；避免因水温问题放大而导致更大的损失和更高的维修费用；解决水温故障，保持发动机在正常的温度范围内运作，可以最大限度地延长使用寿命。

综上所述，解决水温故障对车企及相关企业的价值非常大，车企应该通过改进设计、提高制造工艺以及提供及时故障预警售后服务等方式，尽可能避免和解决水温故障。

对于 A 企业来说，实现发动机水温故障预警具有以下价值：

（1）降低维修成本：发动机水温故障需要进行维修或更换部件，而这些成本通常相对较高。实现水温故障预警可以提前发现问题，及时采取措施进行处理，避免或减少因故障而造成的高昂维修费用，降低企业的维修成本。

（2）提高车辆可靠性：发动机水温异常会导致发动机性能下降，严重时会造成发动机损坏，甚至引发车辆安全事故。实现水温故障预警可以帮助企业及时发现发动机水温异常，及时采取措施进行处理，避免或减少损害，提高车辆的可靠性和安全性。

（3）提升客户满意度：发动机水温故障会影响车辆性能和寿命，导致客户不满意。实现水温故障预警可以帮助企业及时发现问题并进行处理，提高客户的满意度，增强品牌形象。

实现水温故障预警的需求，需要通过实施车联网项目来完成，因为只有通过车联网技术，才能够实时监测车辆的水温信息，并将其传输至云端进行分析和处理，从而实现水温异常的预警和提醒。具体来说，实施车联网项目

需要在车辆中安装传感器和设备来收集车辆的数据，并通过无线通信技术将数据传输至云端进行处理。在云端，可以使用数据分析和人工智能技术来实现水温异常的预测和预警，并通过车载终端或手机应用程序向车主发送提醒和建议。

本文通过收集 A 企业相关数据，标记故障运行数据，利用机器学习算法对数据进行训练和预测，实现发动机水温故障预警。故障预警的应用可以帮助车主及时发现发动机水温异常，从而采取措施进行处理，避免或减少损害。因此，基于机器学习的故障预警在车联网项目风险管理中的应用研究具有重要的意义和价值。

3.2.3 特征选取

在对 A 企业车联网项目团队调研后，本文按照实际业务划分，即按照系统划分，最终选取了 9 类系统，一共 75 个车辆数据特征用来构建故障预测模型。具体系统类别有：空气系统、车身系统、后处理系统、扭矩输出系统、燃油系统、发动机传感器系统、发动机控制系统、变速箱系统、故障诊断系统。

(1) 空气系统

空气系统是通过测量进气量、排气量、进气压力、进气温度、EGR 开度等传感器模拟值的测量，进行一系列计算并输出控制命令，以控制相应的执行器（例如节气门、EGR 阀等），来确保发动机在不同的运行模式下达到设定要求。空气系统中的各个传感器和执行器，都是紧密相连的，它们的运行状态相互影响，因此空气系统的故障可能导致其他系统的异常表现。

空气系统中的进气温度传感器，是用于测量进气温度的一种传感器，如果该传感器出现故障，可能会导致发动机水温异常升高的情况，因为进气温度对于发动机燃烧过程中的空气密度和压力都有着直接影响。如果进气温度传感器的测量值不准确，发动机控制单元将会得到错误的进气温度数据，从而导致错误的气燃比，进而加剧燃烧过程，使发动机产生过热。因此，空气系统和水温故障之间是存在关联的。

空气系统特征具体如下表 3-3 所示。

表 3-3 空气系统特征说明

特征名称	特征说明
AirMod_mEgrPerCyl_f(mg/Hub)	EGR 实际循环废气量(mg/Hub)
AirMod_mfGasThrVlv_f(kg/h)	新鲜进气流量(kg/h)
Air_pIntkVUs(hPa)	增压压力(hPa)
AirCtl_stMon(-)	空气控制管理监控(-)
ASMod_tCmprDs(deg C)	中冷前进气温度 T20(deg C)
ASMod_tEG_mp(deg C)	排气总管 EGR 分流之前温度 T3(deg C)
ASMod_tIntMnfDs(deg C)	进气总管混合点之后温度 T22(deg C)
ChrSet_mEgrDesVal(mg/Hub)	EGR 需求循环废气量(mg/Hub)
EGRVlv_r(%)	EGR 需求开度(%)
EGRVlv_rAct(%)	EGR 实际反馈开度 (H 桥) (%)
ThrVlv_r(%)	节气门需求开度(%)
ThrVlv_rAct(%)	节气门实际关度(%)

(2) 车身系统

车身系统是由整车传感器和执行器构成的系统，用以监测并传输整车行驶各种参数给对应处理器，以实现车辆的控制和管理。空气系统是车身系统的一个重要组成部分，车身系统的整体协调和优化也需要考虑到空气系统的影响，空气系统作为发动机的空气进入量和质量的监测和控制设备，直接影响着发动机的运行状态，进而对发动机水温产生影响。特征如下表 3-4。

表 3-4 车身系统特征说明

特征名称	特征说明
ACCtl_stOut(-)	空调控制状态(-)
Brk_st(-)	刹车开关状态(-)
Clth_st(-)	离合开关状态(-)
EBrk_stLoIdleSwT(-)	发动机制动开关状态(-)
PTOSwt_st(-)	多态开关状态(-)
Com_stRxErc1RtdrTrqMode(-)	缓速器模式(-)
Com_rRxErc1ActRtdrTrqPrc(%)	缓速器实际扭矩百分比(%)
ExhFlpTrbnDs_r(%)	排气制动激活状态(%)
EngBrk_stECRActrDem(-)	缸内制动激活状态(-)
Fan_nFlt(rpm)	风扇实际转速(rpm)
VehV_v(km/h)	车速(km/h)

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/588051017024006027>