

doi:10.11835/j.issn.1000-582X.2025.08.009

引用格式:张莹,袁海兵,何祺,等.基于机器学习的汽车智能座舱告警筛选系统[J].重庆大学学报,2025,48(8): 99-110.



基于机器学习的汽车智能座舱告警筛选系统

张莹¹,袁海兵²,何祺³,姜立标^{4a,5},陈毅锋⁶,陈桥芳^{4b}

(1. 星河智联汽车科技有限公司,广州 511400; 2. 广汽能源科技有限公司,广州 510800; 3. 广汽丰田汽车有限公司,广州 511455; 4. 广州城市理工学院 a. 机械工程学院与机器人学院; b. 工程研究院,广州 510800; 5. 华南理工大学机械与汽车工程学院 广州 510641; 6. 重庆理工大学车辆工程学院,重庆 400054)

摘要:本研究探讨了一种基于机器学习技术的汽车智能座舱告警系统,旨在应对由众多告警源引发的安全风险问题。针对现行系统中告警信息的冗余和分类不精确等缺陷,本文提出了一种结合人工经验筛选法与 CNN 模型的混合筛选策略。具体而言,该策略通过整合来自不同设备的运行状态信息并进行有效分类,利用人工经验减少疑似缺陷信号,同时借助卷积神经网络(convolutional neural network, CNN)模型进行特征提取和精准分类。实验结果显示, CNN 模型在测试集上的分类准确率达到 89.07%,而将 2 种方法综合运用后,对所有原始告警信号的筛选准确率更是高达 99.998%,显著超越了现有 VAS 系统的筛选准确率(90%)。这验证了所提出方法在告警信息筛选方面的高效性和卓越性。未来的研究将着重于增加训练数据量、优化模型参数以及改进文本预处理技术等方面,以期进一步提升系统的整体性能。

关键词:机器学习技术;智能座舱告警;告警源;CNN

中图分类号:U469.72+2

文献标志码:A

文章编号:1000-582X(2025)08-099-12

Machine learning-based intelligent cabin alert filtering system for vehicles

ZHANG Ying¹, YUAN Haibing², HE Qi³, JIANG Libiao^{4a,5}, CHEN Yifeng⁶, CHEN Qiaofang^{4b}

(1. Syncore Autotech Co., Ltd., Guangzhou 511400, P. R. China; 2. GAC Energy Technology Co., LTD., Guangzhou 510800, P. R. China; 3. GAC Toyota Motor Co., LTD., Guangzhou 511455, P. R. China; 4a. School of Mechanical Engineering and Robotics; 4b. Institute of Engineering Research, Guangzhou City University of Technology, Guangzhou 510800, P. R. China; 5. School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510641, P. R. China; 6. School of Vehicle Engineering, Chongqing University of Technology, Chongqing 400054, P. R. China)

Abstract: This study presents a machine learning-based intelligent cabin alert filtering system for vehicles aiming to address safety risks caused by excessive and redundant alarm sources. To overcome limitations in current systems, such as alarm redundancy and inaccurate classifications, a hybrid selection strategy is proposed that

收稿日期:2024-11-02

基金项目:国家自然科学基金(61602345)。

Supported by National Natural Science Foundation of China(61602345).

作者简介:张莹(1981—),男,高级工程师,主要从事智能汽车方向研究,(E-mail)yingzhang@syncore.space。

通信作者:姜立标(1965—),男,副教授,(E-mail)jlb620620@163.com。

combines manual expert filtering with a convolutional neural network (CNN) model. The system integrates operational data from various devices, applying manual heuristics to eliminate likely false signals and employing the CNN model for robust feature extraction and precise classification. Experimental results show that the CNN model achieves a classification accuracy of 89.07% on the test dataset. When combined with manual filtering, the overall selection accuracy of alarm signals reaches 99.998%, significantly surpassing the conventional VAS system (90%). These results validate the proposed method's effectiveness in filtering alarm information. Future research will focus on expanding training datasets, optimizing model parameters, and improving text pre-processing techniques to further enhance the overall system performance.

Keywords: machine learning; intelligent cabin alarms; alarm filtering; CNN

汽车是现代社会重要的交通工具,其安全性受到广泛关注。随着智能交通系统的发展,汽车座舱集成了众多电子系统和信息显示装置,如车载导航、驾驶辅助系统等。这些系统在提供便利的同时,可能导致信息过载^[1]。智能告警系统在座舱人机交互中扮演关键角色,负责及时准确地传递关键安全信息,辅助驾驶员做出正确决策,避免或减轻事故危害。汽车配备多种传感器监测车辆状态和周围环境。例如,轮速传感器、加速度传感器、陀螺仪等实时获取车辆行驶动态信息,如:车速、加速度、转向角度等^[2]。环境感知传感器如:毫米波雷达、摄像头、激光雷达监测车辆周围障碍物、其他车辆及行人的位置与运动状态^[3]。车辆总线数据通过车内通信网络收集来自发动机管理系统、制动系统、电池管理系统(针对电动汽车)等电子控制单元(electronic control unit, ECU)的数据,反映车辆关键部件工作状况,为智能告警系统提供车辆健康状态依据^[4]。

告警系统在汽车座舱生产运行管理中至关重要,及时提醒操作人员和工厂人员异常状态或设备故障。智能监控告警信息是监控设备运行状态的关键手段^[5]。设计一套高效智能预警系统,能提升设备操作安全性和保障车辆安全,实现经济和社会价值。然而,过多的告警源会导致系统小故障触发大量告警信号,而大故障可能无法得到及时发现^[6]。因此,众多研究人员开始探索利用数据挖掘技术提升智能预警系统的工作效率与控制能力。早期的方法主要基于专家经验及特定领域的知识积累,随着监测数据规模的增长及机器学习方法的进步,现代警报系统逐步融合了人工智能技术,借助对语义信息的智能分析与处理手段,构建一系列高效的故障识别模型,为汽车座舱报警功能的优化与开发提供了有力支持^[7]。

告警呈现方式包括:视觉告警、听觉告警和触觉告警。视觉告警传统上通过仪表盘上的指示灯提示车辆故障、系统状态等信息。改进后的抬头显示系统(head up display, HUD)将重要信息投影在挡风玻璃上,减少视线转移。听觉告警包括蜂鸣声和语音提示,适用于多种驾驶场景^[8]。触觉告警通过座椅震动和方向盘震动提示驾驶员,不占用视觉、听觉通道,有效吸引驾驶员注意。驾驶员对告警的响应特性包括:视觉注意力分配、听觉认知负荷和触觉感知敏感度^[9]。视觉告警位置、亮度、颜色和呈现时间影响驾驶员视线转移速度、准确性及回视道路的及时性。听觉告警过多或不合理会增加驾驶员认知负荷,干扰正常驾驶信息处理。触觉告警的敏感度存在个体差异,且在不同驾驶状态下反应不同^[10]。当前告警系统研究存在问题,包括:多源信息融合的精准度与可靠性、告警个性化程度不足和人因工程考虑欠缺。不同传感器在精度、可靠性、数据更新频率等方面存在差异,影响危险预测与告警触发^[11]。现有智能告警系统采用统一告警策略,未充分考虑驾驶员个体差异、驾驶习惯及驾驶任务特殊性。告警系统设计过程中未能全面优化告警与人的交互体验^[12]。

汽车座舱智能报警系统的开发目标是对各类警报信息进行统一管理与有效调度,其功能涵盖警报数据的整合、分析与分级处理,同时涉及机器学习模型的构建、验证及性能评估。此外,系统还需要实现故障信息的主动推送、自动报告生成及人机交互界面的合理布局与设计^[13]。如图1所示,本研究的总体流程围绕系统设计展开,旨在构建一个具备良好适应性、安全性与识别准确性的智能报警机制,实现对潜在风险的提前预警,提高汽车座舱整体的安全水平^[14]。

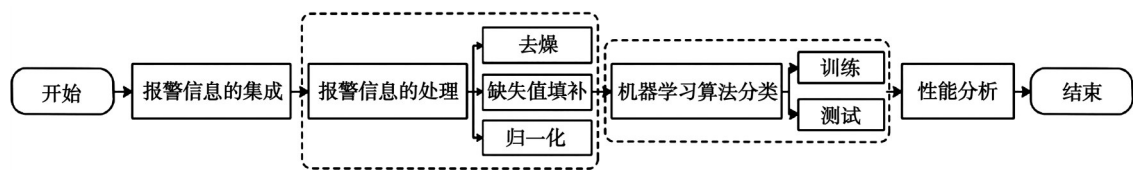


图 1 基于机器学习的汽车座舱智能告警信息分类流程

Fig. 1 Machine learning-based intelligent alert information classification process for vehicle cockpits

智能报警系统中的告警信息整合流程主要涉及对来自计算机监控系统中各类设备运行状态的数据采集与归类。这些数据涵盖稳态、动态及暂态等多种类型的信息,通过统一方式存储至集中式数据库,为后续分析和建模提供基础支持^[15]。此类集成机制有助于提升整体系统的可观测性与可控性,便于工作人员全面掌握设备运行状况。在汽车座舱应用场景中,告警信号依据其影响程度通常被划分为 5 类:异常、事故、越限、变位以及告知。其中,前 3 类具有较高优先级^[16],在系统运行过程中实现快速响应。异常信号直接反映了设备当前的运行状态;事故信号多由硬件故障引发,如开关跳闸等,通过保护装置触发告警,提示维护人员及时介入;越限信号则表示某项参数超出预设的安全范围,属于需要重点关注的预警信息;变位信号主要体现开关类部件的状态变化过程,例如断开或闭合操作,这类信号记录了系统状态的动态演变,也应实时采集并纳入监控体系^[17];告知类信号作为最低级别的告警信息,在设备正常运行时持续反馈状态,起到辅助监测作用。针对不同类型告警信号的特征差异与优先等级,应采取相应的整合策略以提高信息处理效率^[18]。笔者在汽车座舱场景下采用了联邦数据库架构来实现告警数据的集成管理,该架构有效支持来自多种异构设备数据源之间的分布式接入与统一调度^[19]。为进一步提升分类模型的识别精度,还需要对整合后的数据执行一系列预处理步骤,包括:噪声抑制、缺失值填充及数值归一化等关键环节。考虑到智能告警信息具有的高维分布性和来源多样性特点,本研究采用小波阈值去噪法对原始数据进行降噪处理,具体流程包括:数据矩阵转换、二维小波分解、系数阈值调整及信号重构等步骤。经过降噪后的数据再利用 K 近邻算法完成缺失值的填补,最终通过归一化操作形成结构化且标准化的数据集,用于后续模型训练与性能验证^[20]。

汽车智能座舱告警筛选方法包括:1)经验筛选法,利用专家经验和规则库筛选告警,优点是简单易实现、可解释性强,但维护成本高、易误报漏报;2)神经网络经验筛选法,通过神经网络学习历史数据和反馈构建模型,优点是适应性强,但模型解释性差,需大量高质量数据;3)文本向量法:将告警文本转化为向量,用相似度方法筛选,优点是处理文本信息、挖掘语义,但对预处理要求高,难以处理模糊信息。

当下研究不足之处,最主要为现有方法未充分考虑驾驶员个性化需求,缺乏多模态信息融合,且实时性和计算效率有待提升。研究针对以上情况基于机器学习进行个性化实时告警筛选,基于知识图谱的告警信息解释,旨在提高筛选准确性和可靠性,增强可解释性和可信度。

1 汽车座舱告警信息概述

汽车作为现代社会最重要的交通工具之一,其安全性一直备受关注。在智能交通系统蓬勃发展的背景下,汽车座舱内集成了越来越多的电子系统和信息显示装置,如:车载导航、驾驶辅助系统、车辆状态监测系统等。这些系统在为驾驶员提供便利的同时,也带来信息过载的风险。智能告警系统作为座舱人机交互的关键环节,肩负着及时、准确地向驾驶员传递关键安全信息,辅助驾驶员做出正确决策,避免或减轻事故危害的重任^[21]。

原理与含义:智能驾驶辅助系统整合了包括自适应巡航控制、车道保持辅助、自动紧急制动等多项功能。这些系统利用雷达、摄像头等多种传感器监测车辆周围的环境以及自身行驶状态。在检测到潜在危险或系统功能异常时,系统会发出相应告警^[22]。自适应巡航控制告警:自适应巡航控制系统能够自动调整车速维持与前车的安全距离。若前方雷达传感器被遮挡(如积雪或污垢覆盖)或系统故障,无法准确计算与前车的距离,将触发自适应巡航控制告警。此时,系统可能自动停止工作,并提示驾驶员手动控制车速。车道保持辅

助告警:车道保持辅助系统通过摄像头识别车道线,当车辆偏离车道时,系统会轻微转动方向盘辅助车辆回归车道^[23]。若摄像头视野受阻(如雨水、强光干扰)或系统无法正确识别车道线,将发出车道保持辅助告警。这可能导致系统暂时无法正常工作,驾驶员需要更注意车道保持。自动紧急制动系统告警:该系统在检测到即将与前方障碍物碰撞时会自动执行制动。若传感器故障,如雷达信号错误或软件算法问题,可能会触发自动紧急制动系统告警,导致系统误触发制动或在必要时无法正常工作,对行车安全构成威胁^[24]。

在智能汽车领域,汽车设备会在执行动作或出现故障时自动且实时地发出告警信号。这些信号被收集至汽车内的远动机,通过调度数据网传输至调度部门的智能告警系统。该系统用于集中整合其管理范围内各类汽车设备所产生的实时告警信号,依据预设逻辑对信息进行过滤、归类或可视化输出。由于智能汽车中搭载的设备种类繁多,且每条告警记录通常包含较为详细的事件描述,因此,系统整体产生的告警数据规模较大。然而,这些告警信号中的大多数并不代表设备存在缺陷,因此,准确识别缺陷告警信号是关键的研究课题。面对庞大的原始告警数据集,传统的人工经验筛选方式主要基于对历史记录与实际缺陷情况的统计分析,制定相应的信息过滤规则,实现对关键设备故障信号的有效识别^[25]。该方法通常包括3个核心环节:字段划分、关键词提取及筛选策略设计。通过对真实数据样本的深入分析发现,汽车设备常见的异常类型主要包括:未恢复状态、时间同步偏差以及动作频率异常3类,分别对应于关键字匹配法、时间戳对比法以及频次统计法3种筛选机制,如图2所示。

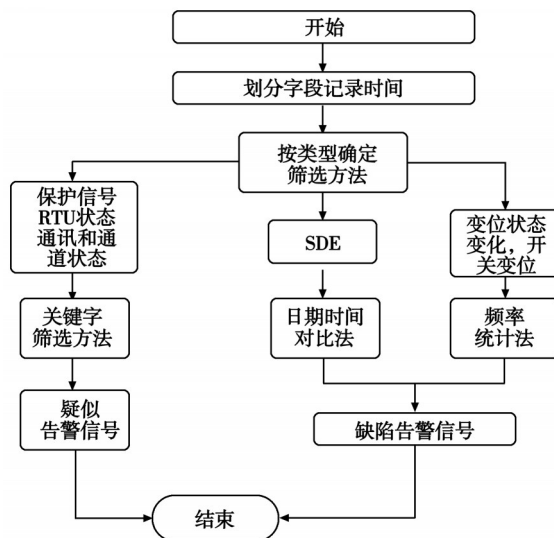


图2 人工经验筛选算法整体流程

Fig. 2 Framework of artificial experience filtering algorithm

本研究通过人工经验筛选策略对告警系统输出的各类信号进行逐项实时解析,旨在甄别出与设备缺陷存在关联的潜在告警信息。相较于原始告警数据,经该方法处理后的告警数量显著降低。通过频率统计与时间序列对比,进一步确认部分疑似缺陷信号是否真实反映设备异常行为,例如装置频繁动作或事件顺序(sequence of event, SOE)存在时间偏差等情况。然而,结合实际缺陷统计数据发现,并非所有通过关键词匹配获得的疑似缺陷告警均为有效缺陷信号,可能包含干扰信息或冗余内容。因此,这些初步筛选结果将作为后续机器学习与自然语言处理技术的输入数据,用于更深层次的特征提取与分类判断。该策略可在大幅压缩待分析告警数量的同时,初步识别部分真实缺陷信号,为基于机器学习的筛选方法提供有效数据基础。

汽车座舱智能告警系统作为保障行车安全的关键技术,在过去几十年取得显著进展,从简单的指示灯告警发展到如今融合视觉、听觉、触觉多种模态的智能告警体系。然而,当前研究仍面临多源信息融合、告警个性化、人因工程优化等诸多挑战^[26]。展望未来,随着新技术的不断涌现与应用,智能告警系统有望实现更精准的危险感知、高度个性化的交互以及沉浸式的告警体验,为驾驶员提供全方位、高效能的安全保障,推动汽车智能化驾驶向更安全、更可靠的方向发展。

2 告警信息机器学习筛选模型

2.1 人工筛选经验法

在对大量原始告警信号应用人工经验筛选方法后,可以识别部分真实的缺陷告警,并从整体数据中提取少量疑似缺陷信号,显著降低需要进一步分析的告警数量。根据实际缺陷统计结果可知,尽管通过关键词匹配所获取的部分告警类型如:保护信号、RTU 状态、通道状态、通讯状态、事故总信号、遥控操作等包含设备异常的相关描述,但其中仍存在一定比例的干扰或冗余信息,非全部属于真实缺陷信号。因此,有必要对这些初步筛选结果进行更深入的甄别处理^[27]。研究所提出的机器学习筛选方法正是针对上述 6 类疑似缺陷告警进行精细化分类的关键步骤。为提升对不确定信号的识别准确性,引入基于机器学习的二次筛选机制。接下来将介绍自然语言处理中涉及的机器学习相关理论,并详细阐述文本分类任务中的主要处理流程,包括:文本预处理、向量化表示、分类模型训练与测试等内容,同时构建适用于告警信号筛选的机器学习分析框架。由于该方法不依赖于时间戳、信号类型或所属场站等结构化字段,仅依据缺陷描述文本内容进行判断。因此,在对信号进行处理时,着重分析其“内容字段”所承载的语义特征。在筛选实施过程中,“漏选信号”指的是本应被判定为存在缺陷的信号却被错误归类为正常样本的情况,此类信号在全部缺陷样本中所占比例称为漏选率;而“误选信号”则表示将原本无缺陷的样本误判为缺陷信号的现象,误选率反映这类错误判断在所有正常样本中所占的比例。

在利用机器学习技术处理自然语言信息时,须按照标准流程进行相应操作。由于机器学习模型仅能处理数值型输入,在将文本信息输入模型前,必须经过一系列预处理步骤,将其转化为可识别的数值形式。研究主要包括 5 个关键环节:训练集与测试集构建、文本清洗与标准化、特征向量化、模型分类训练及性能评估,具体研究流程如图 3 所示。

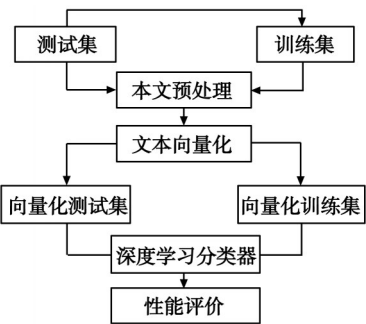


图 3 机器学习筛选算法整体流程
Fig. 3 Framework of deep learning filtering algorithm

- 各部分内容的实现方式如下:
- 1)训练集与测试集构建:告警数据涵盖了 6 种可能存在设备缺陷特征的信号类型,分为训练集和测试集。使用有监督的机器学习模型,通过训练集数据训练模型学习非线性映射,用测试集数据评估模型准确率。模型输入为原始告警信号,输出为二分类结果标签,缺陷信号标记为“1”,非缺陷信号标记为“0”。
 - 2)文本预处理:处理原始数据以适应文本向量化,包括:统一格式、过滤字符、分词和对齐。统一格式涉及将告警信号编码为字符串;过滤字符去除无效字符;分词将句子拆分为词语;对齐确保文本词数一致,不足的用空白词填充。
 - 3)文本向量化:在自然语言处理中,为使机器学习模型能有效识别和分类文本信息,将经过预处理的词语转化为数值型向量形式,完成特征提取与表示转换。该过程的核心目标是将词汇映射为具有语义表达能力的低维实数向量,提升模型对文本信息的处理效果。为便于后续计算与建模,每个词语均被表示为相同维度的数字向量,整段文本则以二维矩阵的形式输入模型中分析。
 - 4)机器学习分类器:本研究采用卷积神经网络模型对文本信息进行特征提取与分类分析,实现告警信号的自动识别与筛选。尽管卷积神经网络常用于图像处理,但在自然语言处理任务中表现出较强的局部特征

捕捉能力。该模型主要包括:输入层、卷积层和输出层,其中卷积层负责自动提取文本的关键语义特征,通过全连接层结合 Softmax 函数完成最终分类决策。作为一种监督学习方法,该模型先使用训练集进行参数学习,再利用测试集评估其分类性能。模型输出结果为二值标签,0 表示当前信号不属于真实缺陷,1 则表示判定为有效缺陷告警。

5)性能评价:基于模型训练过程中的输出结果以及分类器在测试集上的表现,对所构建的机器学习告警信号分类系统进行综合性能评估与优化调整。设定包括:准确率、误选率及多选率等关键指标,验证系统在实际应用中的可行性与有效性。

2.2 神经网络筛选经验法

卷积神经网络(convolutional neural network, CNN)是一种模仿生物神经系统运作机制的计算模型,通过构建多层次的人工神经元模拟复杂的非线性映射关系。该模型由一系列相互连接的节点组成,这些节点通过彼此间的交互作用处理并转换输入数据。CNN 被视为一种拥有高度抽象能力的系统,在众多输入、输出实例中自主学习二者间潜在的相关性。通过对大规模数据集进行反复训练, CNN 依据预测误差调整内部参数,提高识别精度,逼近复杂非线性映射。典型的 CNN 架构通常包括 3 个主要层次:输入层、隐藏层和输出层。在前馈网络结构中,信息从输入层开始,经过一个或多个隐藏层处理后传递到输出层,整个流程为单向流动,没有反馈回路。不同层级之间通过加权连接实现全连接机制,即上一层中的每一个神经元均与下一层中的所有神经元相连,同一层内的神经元之间则无直接连接。该结构保证了模型对输入特征的全面捕捉能力,其整体工作机制如图 4 所示。

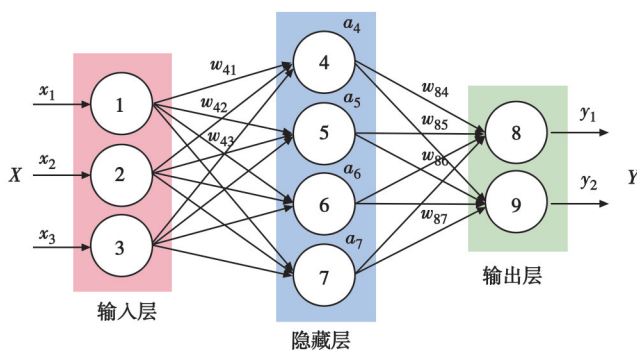


图4 神经网络示意图

Fig. 4 Neural network

2.3 文本向量法

由于告警信息通常由汉字、英文字母及数字混合构成,无法被机器学习模型直接处理,在建模前需要对原始文本进行数字化转换。为使分类器能够有效解析告警内容,应选择合适的向量化策略,将分词后的词语映射为对应的数值表示,并将每条告警视为若干词语的组合,构造成二维矩阵形式,作为卷积神经网络分类器的输入结构。

在自然语言处理领域,常见的文本向量化方法主要包括 2 类:一类是基于统计频率的词袋模型(bag of words, BoW);另一类是基于语义分布的分布式表示方法(distribution representation, DR)。其中,词袋模型是一种较基础且高效的文本编码方式。该方法通过对语料库中出现的所有词汇进行频率统计,按出现频次排序后分配唯一编号。通常高频词获得较小的索引值,低频词对应较大的数值,实现对词语的数字化表示。其优势在于实现简单、计算开销小,但存在缺乏语义表达能力和维度灾难等问题,难以适应复杂语义任务的需求。考虑到本研究中的告警数据具有表述规范、重复性强、语义信息不突出等特点,且大部分文本来源于设备自动生成,对深层语义理解的要求较低,为提高模型运行效率,研究选用词袋模型中的词编码函数(word encoding function, WEF)进行文本向量化处理。该函数依据词语在整个语料库中的出现频率依次赋予编码,具备实现简便、资源占用少等优点,可有效控制特征维度,避免高维稀疏问题。

相比之下,词的分布式表示方法是一种更先进的向量化技术,也被称为词嵌入(word embedding)方法。

它通过数学建模的方式将词语映射为低维稠密的实数向量,较好捕捉词语之间的语义关联性。对于需要深层次语义理解的任务,如情感分析或句法结构识别,此类方法展现出更优的性能表现。主要优势包括:

- 1)词间相似性体现:词向量之间可通过欧氏距离衡量相似程度,语义相近的词语其向量距离更小;
- 2)语义信息丰富:词向量能够捕捉词语之间的深层语义关系,每一维度代表特定的语言特征;
- 3)维度可控:相比传统方法,词向量维度较低,有助于缓解高维带来的计算负担。

当前主流的词向量构建方法主要包括2类:基于统计特征的方法和基于神经语言模型的方法,具体介绍如下:

1)基于统计特征的方法:此类技术通常通过构建词语共现矩阵,结合奇异值分解等数学手段生成对应的词向量。尽管该类方法能够在一定程度上刻画词语之间的相关关系,但由于其生成的向量普遍存在稀疏性高、维度庞大等问题,在实际应用中已较少被采用。

2)基于神经语言模型的方法:该方式依赖于神经语言模型(neural network language model,NNLM)的训练过程,最终输出即为所需要的词向量表示。基本原理是借助上下文信息来预测目标词,获得具有较强语义表达能力的嵌入向量。这类方法在自然语言处理任务中表现出更优的泛化能力和适应性。

3 座舱告警信号实验方案

3.1 实验流程

前文已介绍了基于机器学习技术对智能座舱告警信息进行筛选分析的整体流程及关键技术原理,包括人工经验筛选方法与机器学习筛选方法的协同应用机制^[28]。本章节旨在通过实际告警数据的筛选实验,验证所提方法的有效性与实用性。首先,对采集到的大规模原始告警信息进行系统整理,搭建相应的实验平台。随后,依次采用人工经验筛选与机器学习模型进行多阶段联合过滤,获取实验结果,并对相关算法与模型的性能表现进行评估,验证该方法在实际场景中的可行性,为后续研究结论的形成提供数据支撑。

实验所用数据来源于某地市交通局安全3区车辆报警系统(vehicle alerting system,VAS)镜像服务器的历史数据库,完整记录了1年内积累的所有车辆告警信息。同时,各区授权4S店提供的年度设备缺陷统计表作为分类结果的辅助依据。据统计,全年共采集告警信号约 $5\,571\times10^4$ 条,主要类型包括:保护信号、测量超限以及事件顺序记录(SOE)等。运维人员记录的缺陷约550条,主要也是保护信号类。初步筛选采用人工经验和关键字方法,选出疑似缺陷信号。这些信号再经过机器学习方法筛选,得到关键指标如准确率、漏选率等,并得出结论。由于缺陷多与保护信号相关,机器学习方法成为筛选重点,综合性能指标基于整体筛选结果得出,具体筛选步骤如图5所示。

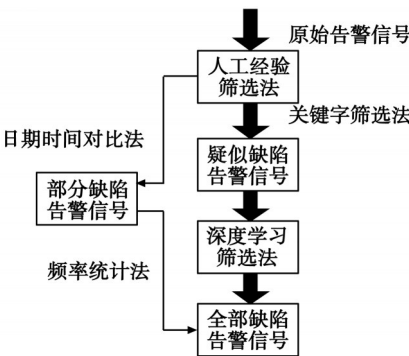


图5 筛选方法整体流程
Fig. 5 Framework of the filtering method

原始告警信号按照时间顺序逐一接收,首先通过人工经验筛选方法进行初步处理,将筛选结果作为输入数据送入机器学习筛选模型中进行深入分析。人工经验筛选环节部署在一台小型服务器上,使用Microsoft SQL Server数据库执行数据校验与管理任务。实验中收集的所有原始告警信息被整合导入数据库,根据

VAS系统记录的时间戳排序后,逐条输入到筛选算法中处理。该阶段的人工经验筛选算法采用SQL语言编写,因其简洁语法和高效执行特性,特别适合于大规模数据集的快速处理。机器学习筛选部分则运行在标准个人电脑上,仅利用CPU完成模型训练与推理计算。筛选算法由MATLAB R2018b实现,涵盖从告警数据输入、模型训练到筛选输出的全过程,生成可视化图表以辅助结果分析。为了验证此方法的有效性,研究对所有历史告警数据进行集中测试。在实际应用场景中,基于预先训练好的机器学习模型对新产生的告警信号实施实时接收与在线分析,提高系统的响应速度及缺陷识别能力^[29]。

3.2 实验过程

对全部原始告警信号实施预处理操作,包括将文本内容统一转换为标准字符串格式,采用文档词语标记工具对其中的中文部分进行分词处理。该工具具备识别常见汉语词汇的能力,实现较为准确的词语切分效果。在完成分词后,进一步将所有英文字符中的大写形式统一转换为小写,并清除文本中包含的各类标点符号。通过上述一系列预处理步骤,有效统一输入数据的格式规范,降低无关信息对后续分析造成的干扰,同时为接下来的文本向量化表示打下良好基础。使用文档词语标记函数前,先构建包含常用词汇的汉语词典作为参考依据。在此基础上,对告警信息中的中文语句进行词语切分与标注,同时具备识别复杂结构的能力,如网址、表情符号及哈希标签等内容。通过绘制词语云图可观察到,该方法在常见词汇识别方面表现良好。在完成分词后,还要进行文本长度标准化处理。为此,首先统计每条告警信号中所含词语的数量,如图6所示,大部分告警信息的词数较少,最长文本共包含18个词语。为了在保留完整信息的同时避免输入维度过高,设定统一输入长度为18,即以最长告警的词数为基准。对于词数不足的样本,在其文本左侧补充占位词,实现输入数据的尺寸对齐,使所有样本在输入机器学习模型时保持一致格式。在完成前述预处理流程后,紧接着构建基于卷积神经网络(CNN)的分类型,并以处理后的告警信号数据及其对应类别标签作为输入,进行训练与验证过程。在模型输入前的关键准备阶段,对文本信息进行向量化表示,使其能够被神经网络有效识别和处理。

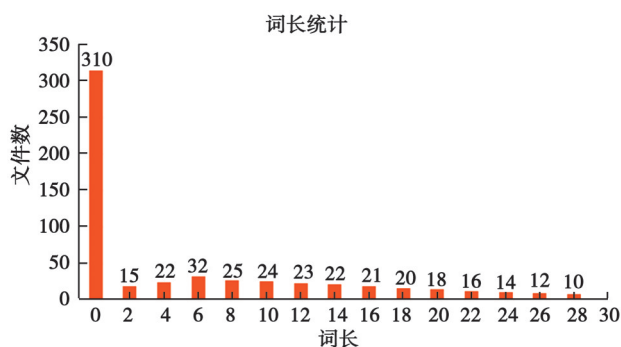


图6 告警信号词长统计

Fig. 6 Alarm signal word length statistics

考虑到本文所处理的告警信息来源于车载系统,其文本内容通常具有结构清晰、用词规范、语义直观等特点,属于表达形式较为简洁的短文本类型,一般不涉及复杂的语义理解。为兼顾计算效率与模型实用性,降低整体复杂度,研究采用词袋模型中的词编码函数对文本进行数值化转换。该方法可将每个词语映射为一个固定长度的实数向量,确保输入格式的一致性,同时满足模型高效运算的需求。

3.3 结果分析

通过多轮实验对比,最终确定了最优的卷积神经网络(CNN)模型参数,如表1所示。研究所使用的数据集涵盖了全年采集的缺陷告警信息,样本规模较大,具有较强的代表性与统计可信度。首先,采用人工经验筛选方法对原始告警信号进行初步过滤,有效减少疑似缺陷信号的数量,使疑似缺陷信号的比例由原来的72.41%下降至53.31%;随后,利用训练完成的CNN模型对剩余告警信息进行精细化二次筛选,实现较高的分类精度。实验结果表明,所构建的CNN模型具备良好的学习能力,预测性能优于传统方法。图7展示了3

种不同算法的测试准确率表现,其中,卷积神经网络的分类准确率达到 89.07%,显示出较好的识别能力。结合人工经验筛选与神经网络的综合方法,其整体告警信号的筛选准确率达到 99.998%,显著高于 VAS 系统所具备的 90% 准确率,验证了本方法在告警信息处理方面的有效性与优越性,适用于实际交通设备监测场景。图 8 展示了各算法在训练过程中的损失变化趋势,总体来看 3 类算法的训练损失均维持在较低水平,表明模型收敛状态良好。

表 1 CNN 机器学习参数
Table 1 CNN machine learning parameters

参数名称	参数值
模型层数	4.000
词向量维度数	25.000
隐藏神经元数	35.000
梯度阈值	2.000
初始学习率	0.004
迭代次数	150.000
优化算法	Adam
分类函数	Soft Max

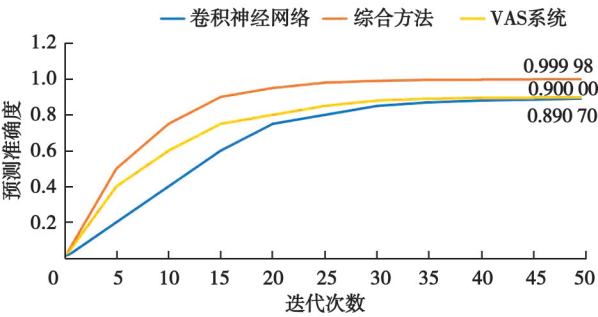


图 7 预测准确度
Fig. 7 Prediction accuracy

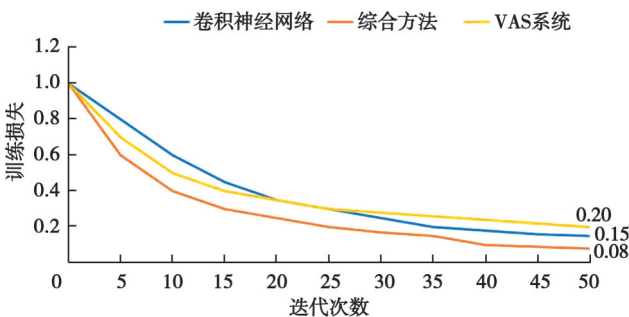


图 8 训练损失
Fig. 8 Training loss

从算法压缩效率角度分析,如图 9 所示,各模型的压缩比均控制在 0.3 以内,表明其资源消耗较低,具备良好的计算轻量化特性,适合实际部署与运行。此外,研究使用机器学习筛选方法对测试集样本进行了分类统计分析,其置信度分布情况详见图 10。其中属于 0.6 类别的样本中,识别出缺陷告警 104 项(时间维度),多维属性告警 152 项。在置信度为 0.7 的类别中,共有 78 条时间维度告警和 124 条多维属性告警。在置信度为 0.8 的类别中,有 58 条时间维度告警和 83 条多维属性告警。在置信度为 0.9 的类别中,有 36 条时间维度告警和 61 条多维属性告警。而在置信度为 1 的类别中,仅有 9 条时间维度告警和 44 条多维属性告警。随着置信度的提高,数据量逐渐减少。这表明在高置信度下,模型对数据的筛选更加严格,只有少数数据能够达到较高的置信度。

对分类结果中出现误判的情况进行分析,主要存在以下几个方面的原因:

1) 部分用于训练与验证的缺陷信息来源于汽车记录系统而非 VAS 告警系统,2 类系统的描述方式存在一定差异。例如:红绿灯同亮消缺、名称与后台不对应、后台机、挡位显示不正确等术语在原始告警信号中无法找到明确对应的表达,导致样本标签与实际输入信息不匹配,影响数据质量,降低模型的学习效果;

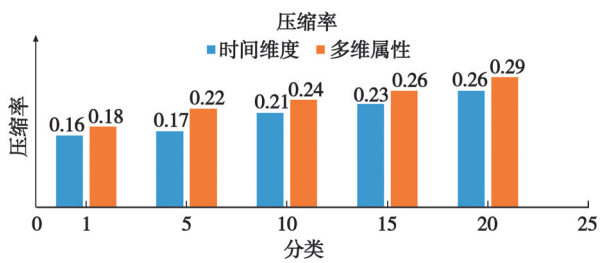


图 9 压缩率
Fig. 9 Compression ratio

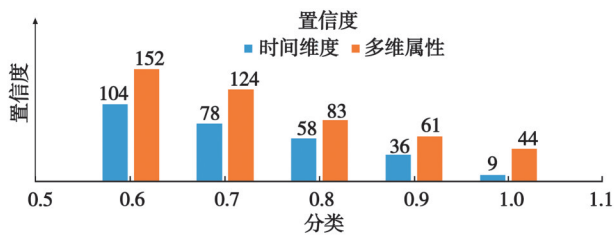


图 10 置信度
Fig. 10 Confidence level

2)在实际运维过程中,存在大量未被记录或遗漏的真实缺陷情况,使训练样本中缺陷样本(正类)比例偏低,分类器难以准确学习关键特征,造成模型泛化能力受限;

3)部分告警信号本身语义相近,表述高度相似但类别归属不同,如某些保护信号之间仅存在细微差异,却归入不同类别,这对模型的区分能力提出较高要求;

4)全年真实缺陷数量相对较少,导致可用于训练的正样本规模偏小,整体数据集中有效样本不足,影响机器学习模型训练的充分性与稳定性。

研究所提出的基于机器学习技术的智能座舱告警信息分析方法在后续仍有进一步改进空间,主要可从以下几个方向展开研究:1)扩大卷积神经网络(CNN)训练样本的数量,以提升模型整体识别精度;2)提高原始告警数据的质量与准确性,确保其与实际缺陷信息保持一致;3)深入探索 CNN 网络结构中的最优参数组合,增强模型泛化能力避免过拟合现象;4)尝试引入其他类型的机器学习分类器,如改进型 CNN 模型,用于提升告警信号分类效果;5)优化文本预处理流程,采用更高效的分词策略及词向量生成方式,提升特征表达能力;6)持续完善人工经验筛选机制,提升输入数据质量并有效过滤冗余告警。

研究所采用的词编码函数算法依据文本中所有唯一词语的出现频率进行数值映射。具体而言,在完成词语频率统计后,按照频率由高到低进行排序,并为每个词语分配相应的编码值。高频词汇赋予较小数值,低频词汇对应较大数值,实现对所有词语的唯一编号转换,最终构建对应的数字向量形式,完成文本向量化操作。

在实施该算法时,首先对所有文本语料中的词语进行提取和统计,共识别出 1 086 个不重复词语。随后,根据各词频分布情况进行降序排列,依次赋予编码 1~1 086。对于长度对齐过程中使用的填充词,则统一赋值为 0,完成全部告警信息的向量化表示。完成上述文本预处理之后,进入卷积神经网络分类器的训练与测试阶段。通过使用训练集数据多次迭代训练获得最优模型参数,再利用测试集对模型性能进行评估,最终得出分类准确率指标,作为模型有效性的重要参考依据。此外,在系统部署与运行成本方面,本文结合系统熵理论计算了各模块的单位算力成本,结果详见表 2。单位成本反映了获取单位计算资源所需要付出的成本代价。从计算结果来看,告警筛选系统的单位成本低于常规系统水平,尤其 CNN 方法的单位成本显著低于传统特征分类方法,表明其具备更高的性价比优势。因此,在实际应用中优先采用 CNN 模型输出结果,可在控制成本的同时实现良好的分析效果。

表 2 单位成本
Table 2 Unit cost

类型	单位成本系数
一般系统	3.00
常规告警系统	2.61
CNN 方法	1.86
特征告警方法	1.53

4 结 论

在智能汽车座舱系统中,对危险告警的识别主要依赖于座舱告警系统所提供的信息。然而,在实际运行过程中,系统产生的告警信号种类繁多,其中包含大量与设备缺陷无关的冗余信号,这为工作人员准确识别关键告警带来较大困难。针对上述问题,研究提出一种融合人工经验筛选与机器学习方法的告警信息分析策略。该方法通过双重筛选机制,实现对原始告警信号的有效分类,能够精准识别出与设备故障相关的异常信号。主要研究内容与结论如下:

1)单位成本是指获取单位计算资源所投入的算力代价。根据系统熵方法计算得出的结果显示,本研究构建的告警筛选系统的单位成本显著低于传统系统水平。尤其在模型输出阶段,卷积神经网络(CNN)所对应的单位成本明显优于传统特征分类方法。因此,在实际部署中优先采用CNN模型的输出结果,有助于在控制成本的同时实现良好的识别效果。

2)在第1阶段筛选中,采用人工经验法成功识别并过滤掉部分非缺陷告警信号,使疑似缺陷信号的比例由原来的72.41%下降至53.31%。随后,在第2阶段引入基于CNN的机器学习筛选方法进行精细化分类。经过训练与测试,模型在测试集上的分类准确率达到89.07%,表现出较强的判别能力。综合2种筛选方式的结果,最终对全部原始告警信号的整体筛选准确率高达99.998%,远高于现有VAS系统约90%的准确率。实验结果充分验证了该方法在实际应用中的有效性与优越性。

3)模型压缩比均控制在0.3以内,表明计算资源占用率低,具备工程化部署的可行性。通过置信度分级筛选机制,样本数量随置信度阈值的提升呈现显著衰减趋势,表明模型在高置信度区间对输入数据的判别标准更为严格,有效过滤了低可信度样本。从工业应用角度而言,为告警信号处理提供了高效、实用的解决方案。

参考文献

- [1] Idlahcen F, Idri A, Goceri E. Exploring data mining and machine learning in gynecologic oncology[J]. Artificial Intelligence Review, 2024, 57(2): 7-12.
- [2] Leonard B. Well prepared: How trichome polymorphism creates an early-warning system against herbivory[J]. The Plant Cell, 2024, 4(12): 6-8.
- [3] 秦雅琴,夏玉兰,钱正富,等. 微观轨迹信息驱动的Bi-LSTM合流区车速预测[J]. 重庆大学学报, 2023, 46(4): 120-128.
Qin Y Q, Xia Y L, Qian Z F, et al. Vehicle speed prediction in Bi-LSTM confluence area driven by micro-trajectory information[J]. Journal of Chongqing University, 2023, 46(4): 120-128. (in chinese)
- [4] 余伟,曾孝平. 麻醉术中绝对血容量不足临床监测告警系统的设计[J]. 重庆大学学报, 2011, 21(8): 43-47.
Yu W, Zeng X P. Design of clinical monitoring alarm system for absolute blood volume deficiency during anesthesia[J]. Journal of Chongqing University, 2011, 21(8): 43-47. (in chinese)
- [5] 高彦,傅春耘,杨忠,等. 基于改进粒子滤波算法的车速估计[J]. 重庆大学学报, 2024, 47(3): 44-52.
Gao Y, Fu C Y, Yang Z, et al. Vehicle speed estimation based on improved particle filter algorithm[J]. Journal of Chongqing University, 2024, 47(3): 44-52. (in chinese)
- [6] 冯新扬,张巧荣,李庆勇. 基于改进型深度网络数据融合的滚动轴承故障识别[J]. 重庆大学学报, 2019, 42(2): 52-62.
Feng X Y, Zhang Q R, Li Q Y. Rolling bearing fault identification based on improved deep network data fusion[J]. Journal of Chongqing University, 2019, 42(2): 52-62. (in chinese)
- [7] 王博文,王景升,王统一,等. 基于长短时记忆网络的Encoder-Decoder多步交通流预测模型[J]. 重庆大学学报, 2021, 44(11): 71-80.
Wang B, Wang J S, Wang T Y, et al. Multi-step traffic flow prediction model based on encoder-decoder with long short-term memory networks[J]. Journal of Chongqing University, 2021, 44(11): 71-80. (in chinese)
- [8] Wang J, Wang Z, Chen M, et al. An interpretable artificial intelligence system for detecting risk factors of gastroesophageal variceal bleeding[J]. Npj Digital Medicine, 2022, 5(1): 114-121.
- [9] Wang Y, Jiménez C, Quiroz M, et al. Tsunami early warning system using offshore tsunameters in Peru[J]. Ocean Engineering,

- 2023, 18(3): 21-31.
- [10] Najafi H , Shrestha P K , Rakovec O, et al.High-resolution impact-based early warning system for riverine flooding[J]. Nature Communications, 2024, 15(1): 4-6.
- [11] Eltoumy K A , Liang X .Uncertainty-aware structural damage warning system using deep variational composite neural networks [J]. Earthquake engineering & structural dynamics, 2023, 3(11): 9-13.
- [12] Braun E J, Singh S, Penlesky A C, et al.Nursing implications of an early warning system implemented to reduce adverse events: a qualitative study[J]. BMJ quality & safety, 2022, 31(10): 716-724.
- [13] Subissi L,Von Gottberg A, Thukral L, et al.An early warning system for emerging SARS-CoV-2 variants[J]. Nature Medicine, 2022, 7(4): 9-11.
- [14] Alonso A . Leak recognition system provides early warning of leaks[J]. Welding Journal, 2023, 6(3): 10-13.
- [15] Murray J, Crowell B, Murray M, et al. Incorporation of real-time earthquake magnitudes estimated via peak ground displacement scaling in the shake alert earthquake early warning system[J]. Bulletin of the Seismological Society of America, 2023, 7(4): 8-11.
- [16] Sievers B L, Siegers J Y, Cadènes J M, et al. Smart markets: harnessing the potential of new technologies for endemic and emerging infectious disease surveillance in traditional food markets[J]. Journal of Virology, 2024, 98(2): 5-9.
- [17] Yang L , Wang Z , Zhang S , et al. Neuromorphic gustatory system with salt-taste perception, information processing, and excessive-intake warning capabilities[J]. Nano letters, 2022, 4(3): 4-7.
- [18] Roberts R D, Goodman S J, Wilson J W, et al. Taking the Highway to save lives on lake victoria[J]. Bulletin of the American Meteorological Society, 2022,7(2): 103-106.
- [19] Tilman A R, Krueger E H, Mcmanus L C, et al. Maintaining human well being as social-environmental systems undergo regime shifts[J]. Ecological Economics, 2024, 4(1), 221:7-10.
- [20] Santiago I D, Plomaritis T A, Avalos D, et al. Comparison of wave overtopping estimation models for urban beaches. Towards an early warning system on the Basque coast[J]. Science of the Total Environment, 2024, 3(2): 10-12.
- [21] Boza J M, Amirali A, Williams S L, et al.Evaluation of a field deployable, high-throughput RT-LAMP device as an early warning system for COVID-19 through SARS-CoV-2 measurements in wastewater[J]. Science of the Total Environment, 2024, 944: 1-11.
- [22] Francisco A P, Silvia G M, Otero M J, et al. Explainable cognitive decline detection in free dialogues with a Machine Learning approach based on pre-trained Large Language Models[J]. Applied Intelligence, 2024, 54(24): 12613-12628.
- [23] He Y, Geng Y, Li M Z. A machine learning approach and numerical investigation for intelligent forecasting of entropy generation rate inside a turbulator-inserted solar collector tube[J]. Engineering analysis with boundary elements, 2024, 158(1): 375-384.
- [24] Song Z, Cao S, Yang H. An interpretable framework for modeling global solar radiation using tree-based ensemble machine learning and Shapley additive explanations methods[J]. Applied Energy, 2024, 364(2): 9-12.
- [25] Guo Z, Deng J, Tang Z Y. A hybrid method of combination probability and machine learning for Chinese geological text segmentation[J]. Computers & geosciences, 2024, 183(1): 1-11.
- [26] Ancona R M, Cooper B P, Randi F, et al.Machine learning classification of new firearm injury encounters in the St Louis region: 2010-2020[J]. Journal of the American Medical Informatics Association, 2024(10): 9-11.
- [27] Li Z, Han M, Vo D N, et al.Machine learning-based input-augmented Koopman modeling and predictive control of nonlinear processes[J].Computers and Chemical Engineering, 2024, 191(13): 7-8.
- [28] Yago D, Sal-Anglada G, Roca D, et al.Machine learning in solid mechanics: application to acoustic metamaterial design[J]. International Journal for Numerical Methods in Engineering, 2024, 125(14): 7-10.
- [29] Li S, Zhang C, Wang X, et al. A machine learning methodology for investigating the liquid-liquid transition of hydrogen under high-pressure[J]. International Journal of Modern Physics C, 2024, 35(12): 4-6.

(编辑 侯 湘)