

DOI:10.19344/j.cnki.issn1671-5276.2026.01.056

基于随机森林算法的电力营销现场工作票问题 分类异常检测研究

尤佳¹, 夏泽举¹, 周永刚¹, 丁建顺¹, 汪亮²

(1. 国网安徽电力有限公司, 安徽 合肥 210000; 2. 南瑞科技有限公司, 江苏 南京 230000)

摘 要:随着电力市场的不断发展和电力营销现场工作票数量的急剧增加,对工作票问题进行准确分类和异常检测成为提高工作效率和保障服务质量的关键。为此提出了基于随机森林算法的电力营销现场工作票问题分类与异常检测方法。通过对历史工作票数据进行预处理,构建电力营销现场工作票数据集。通过自然语言语义分析对工作票问题进行分类,利用随机森林算法的特性,通过计算工作票样本的异常得分,识别出可能存在问题的异常工作票。具体来说,异常得分是基于样本在随机森林算法中的不一致性来计算的,得分较高的样本被认为是异常的。实验结果表明:基于随机森林算法的分类与异常检测方法能够有效地对工作票问题进行分类,并准确地检测出异常工作票,提高了问题处理的效率和准确性为电力营销现场工作票的管理提供了一种新的技术手段,有助于提升电力公司的服务质量和运营效率。

关键词:电力营销;工作票;随机森林算法;问题分类;异常检测

中图分类号:TP274 **文献标志码:**A **文章编号:**1671-5276(2026)01-0299-05

Research on Anomaly Detection of Power Marketing Site WorkOrder Problem Classification Based on Random Forest Algorithm

YOU Jia¹, XIA Zeju¹, ZHOU Yonggang¹, DING Jianshun¹, WANG Liang²

(1. Anhui Electric Power Cooperation, Hefei 210000, China; 2. NARI Technology Co., Ltd., Nanjing 230000, China)

Abstract: As accurate classification and anomaly detection of work order issues become the key to improving work efficiency and ensuring service quality with the development of the electricity market and the sharp increase in the number of work orders in electricity marketing sites, a classification and anomaly detection method for on-site work order problems in power marketing based on random forest algorithm is proposed. This method preprocesses historical work order data to construct a dataset of on-site work orders for power marketing. Natural language semantic analysis is conducted to classify work order issues. By utilizing the characteristics of random forests and calculating the anomaly scores of work order samples, potential problematic abnormal work orders are identified. Specifically, the anomaly score is calculated based on the inconsistency of the samples in the random forest, and samples with higher scores are considered abnormal. The experimental results indicate that the classification and anomaly detection method based on random forest can effectively classify work order problems and accurately detect abnormal work orders, improving the efficiency and accuracy of problem handling. And the proposed method provides a new technological means for the management of on-site work orders in power marketing, conducive to the improvement of service quality and operational efficiency of power companies.

Keywords: electricity marketing; work ticket; random forest; problem classification; anomaly detection

0 引言

随着电力系统规模的不断扩大和智能化程度的提升,工作票中的问题也日益复杂多样,异常事件的发生频率也在逐渐增加。这些异常事件可能源于人为操作失误、设备故障或外部环境影响,若未能及时发现和处理,将可能导致电力设备的损

坏、系统运行的不稳定,甚至造成整个电力系统的崩溃^[1]。为了有效应对这一挑战,异常检测技术在电力营销现场工作票管理中的应用显得尤为重要。基于改进支持向量机(support vector machine, SVM)算法的电力工程异常数据检测方法,通过优化 SVM 算法,提高对电力工程数据中异常值的识别能力,为电力系统的稳定运行提供

基金项目:国家电网公司总部科技项目(5108-202218280A-2-255-XG)

第一作者简介:尤佳(1981—),女,安徽合肥人,高级工程师,本科,研究方向为移动作业技术、电力营销计量采集业务, qyj121588@163.com。

了有力保障。但 SVM 算法本身对于大规模数据集的处理能力有限,且在高维特征空间中可能面临过拟合的风险,从而影响检测精度^[2]。而基于机器学习的电力系统网络异常检测与分类方法,利用机器学习模型的强大学习能力,实现了对电力系统网络异常的有效检测和分类,为电力系统的故障排查和预警提供有力支持。但模型的性能很大程度上依赖于训练数据的质量和数量。如果训练数据中存在噪声或不平衡现象,可能会导致模型泛化能力下降,从而影响异常检测的精度^[3]。为了提升异常检测精度,本文研究基于随机森林算法的电力营销现场工作票问题分类异常检测方法。随机森林算法是一种集成学习方法,通过构建多个决策树进行预测,并综合这些决策树的结果来提高准确性。在异常检测中,随机森林算法能够通过对数据集进行训练,识别出与正常样本不同的异常值,从而实现异常检测的目的。

1 历史工作票数据预处理

在进行基于随机森林算法的电力营销现场工作票问题分类异常检测之前,对历史工作票数据进行预处理。电力营销现场工作票数据主要涵盖工作票 ID、操作类型、设备状态、审批流程、申请时间、作业地点等关键信息字段。这些字段共同构成了工作票的数据结构,并详细记录了电力营销现场的实际作业情况。

经过一系列的处理,如数据清洗(包括处理缺失值、格式统一等)、数据缩放(将数据缩放到0~1的范围)以及降维处理(通过数学变换将高维数据映射到低维空间),将得到一个高质量、结构化的数据集。

在构建电力作业异常行为数据集的过程中,鉴于电力作业的专业性和安全性要求,本文结合真实带电作业场景与模拟不带电作业场景,以确保数据的全面性和真实性^[4]。此时,由于数据集往往存在不规范的问题,设计将数据缩放到一个特定的范围(如0~1),平衡不同工作票数据在量级上的不一致性,可用式(1)加以描述。

$$X_{\text{Dorm}} = \frac{X - \mu}{\sigma} \quad (1)$$

式中: X 为原始电力营销现场工作票数据值(如操作类型的数值编码、设备状态的数值表示等); μ 为该特征在所有工作票数据中的均值; σ 为标准差; X_{Dorm} 为标准化后的工作票数据值^[5]。

在处理高维数据时,通过数学变换,将高维空

间中的工作票数据点映射到低维度的子空间中,从而大幅提高样本密度,简化距离计算,相关表达式为

$$X_{\text{scaled}} = \frac{X_{\text{Dorm}} - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

式中: $X_{\min}$ 和 $X_{\max}$ 分别为各种电力营销现场工作票数据的最小值和最大值^[6]。综上所述,构建一个高质量、规范化的电力营销现场工作票数据集,为后续模型训练和异常检测提供支持。

2 工作票问题分类

在数据预处理的基础上,进一步对电力营销现场工作票问题进行分类。通过对历史数据的分析,总结出信息缺失、审批流程错误、重复申请、填写不规范、逾期未闭环(即逾期未处理)以及其他异常等常见的问题类型,并为每种类型分配唯一的标签^[7]。

在应对工作票问题的分类挑战时,K-means 算法通过循环迭代的方式,将数据巧妙地划分为 K 个独立的簇,而每一个簇都对应着一种特定的异常类型。算法会精确地计算出每个用户数据点到所有簇中心的距离,并依据这些距离信息,将数据点准确地分配到距离最近的簇中,从而圆满实现数据的分类目标^[8]。

假设历史工作票数据集 $X = \{x_1, x_2, \dots, x_m\}$,其中 x_j 是工作票数据的特征向量^[9]。随机选择 K 个工作票数据点作为初始簇中心,记为 $\{\alpha_1, \alpha_2, \dots, \alpha_K\}$ ^[10]。对于每一次迭代,计算每个用户数据点到所有簇中心的距离,并将用户数据点分配到最近的簇中:

$$\text{text}C_{\text{luster}}(x_j) = \arg \min_k \|x_j - \alpha_k\|^2 \quad (3)$$

式中 $\|\cdot\|^2$ 为电力营销现场工作票数据点到簇中心的欧几里得距离。更新簇中心为簇内所有数据点的均值

$$m_k = \frac{1}{|C_{\text{luster},k}|} \sum_{x_j \in C_{\text{luster},k}} x_j \quad (4)$$

式中 $|C_{\text{luster},k}|$ 为簇 k 中的电力营销现场工作票数据点数量。当簇中心不再发生显著变化或达到预设的迭代次数时,迭代终止^[11]。最终,每个数据点都被分配到了一个簇中,这个簇代表了该数据的异常类型。

3 利用随机森林算法计算工作票异常得分

建立工作票问题的分类标签后利用随机森林

算法来计算每张电力营销现场工作票的异常得分。随机森林算法是一种集成学习方法,通过构建多个决策树来提高模型的准确性^[12]。在计算异常得分时,模型会根据训练数据中的经验来评估该工作票的异常程度,并给出一个相应的得分。

在计算异常得分时,首先将每张工作票的特征向量(包含工作票 ID、操作类型、设备状态、审批状态等关键信息)输入到随机森林算法模型中。这些特征向量包含工作票中的关键信息,如操作类型、设备状态等,它们共同构成评估工作票异常程度的基础^[13]。随机森林算法模型在接收到特征向量后,根据训练数据中的经验来评估该工作票的异常程度。这一过程涉及到多个决策树的共同决策,每个决策树都会根据自己的分类规则和训练经验来给出该工作票属于某一类别的概率或得分^[14]。最终,随机森林算法模型会综合多个决策树的输出结果,给出一个反映工作票异常程度的总得分。随机森林算法计算异常得分的过程如式(5)所示。

$$S(x) = \frac{1}{T} \sum_{j=1}^T s_j(x) \quad (5)$$

式中: S 为电力营销现场工作票的异常得分; T 为决策树组成棵数^[15]。这个得分是一个量化指标,它反映了工作票与正常样本之间的偏离程度^[16]。得分越高,表示工作票与正常样本的差异越大,存在异常的可能性也就越大。

4 识别异常工作票类型

根据随机森林算法计算出的异常得分来识别异常电力营销现场工作票。这一步骤通常涉及设置一个阈值,当工作票的异常得分超过该阈值时,即确定该工作票的异常类型。

在识别异常工作票的过程中,首先利用鲸鱼算法进行异常检测流程优化。在随机探索阶段,面对目标位置未知的挑战,算法模拟鲸鱼在海洋中的随机游动,种群中的每个个体都会随机选定一个同伴的位置作为临时目标,并调整自身位置向其靠近^[17]。图 1 展示了鲸鱼优化算法的基本流程,通过这一流程,可以更加高效地识别和处理异常工作票,进一步提高电力运维的效率和安全性。

在得到最终的分数后,为了更直观地展示异常工作票的识别结果,制定如表 1 所示的工作票问题异常划分表。

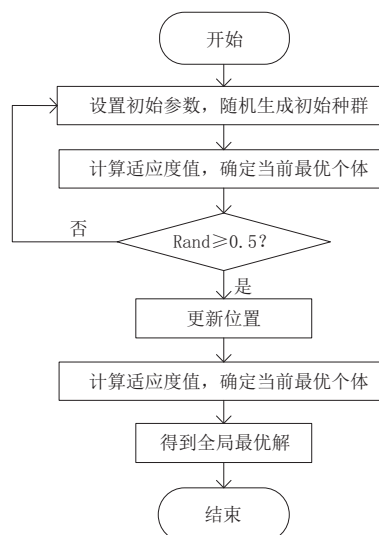


图 1 鲸鱼优化算法

表 1 工作票问题异常划分

序号	工作票问题类型	得分区间	异常阈值
1	信息缺失	0.0~0.3	0.30
2	审批流程错误	0.3~0.7	0.60
3	重复申请	0.2~0.5	0.45
4	填写不规范	0.1~0.4	0.35
5	逾期未处理	0.4~0.8	0.70
6	其他异常	0.2~0.6	0.50

根据随机森林模型计算得出的异常得分,将工作票问题划分为不同的得分区间。得分越高,表示工作票存在该类型问题的可能性越大。同时根据得分区间的分布情况,设定每个问题类型的异常阈值^[18]。当一个工作票的异常得分高于对应问题类型的异常阈值时,可以认为该工作票存在该问题,并需要进一步审查或处理。

5 实验

5.1 实验准备

国网某供电公司长期面临电力营销现场工作票管理复杂、问题多样的挑战。为了提升工作票管理的效率和准确性,该公司决定引入基于随机森林的异常检测算法,对工作票进行智能化分类和异常检测^[19]。

为了实现这一目标,搭建一个适合算法运行和实验验证的环境。如表 2 所示。

表 2 实验环境搭建

设备名称	具体型号	设备用途	使用数量
服务器	Dell PowerEdge R740	数据处理和模型训练的主要计算平台	2 台
高性能计算集群	Lenovo ThinkSystem SR650	加速随机森林模型的训练和预测过程	1 套(含多个节点)
数据存储设备	Huawei OceanStor Dorado 系列全闪存存储	存储实验所需的大量历史工作票数据	1 台
随机森林算法实现软件	scikit-learn (版本 0.24 及以上)	实现随机森林算法,进行工作票问题分类和异常检测	1 套
数据库管理系统	MySQL (版本 8.0 及以上)	存储和管理预处理后的工作票数据	1 套

这一环境不仅支持大规模数据的快速处理,还具备强大的计算能力,能够满足复杂模型的训练需求^[20]。

5.2 实验数据

收集该公司近年来大量的电力营销现场工作票数据,如表 3 所示。利用这些数据对随机森林算法进行了训练和调优,使其能够准确识别出工作票中的异常问题。这一数据集将涵盖电力营销现场工作票中可能出现的各类问题,从而为后续的算法训练和测试提供有力的数据支撑。

表 3 电力营销现场工作票问题数据集

工作票问题类型	训练集(样本数)	测试集(样本数)
信息缺失	5 000	1 000
重复申请	3 000	600
审批流程错误	4 000	800
填写不规范	6 000	1 200
逾期未处理	7 000	1 400

5.3 实验结果与分析

为了全面评估本文算法的异常检测能力,选取 3 种方法进行对比实验,并记录各自的异常检测率。将基于改进 SVM 算法的电力工程异常数据检测方法设计视为方法 1,基于机器学习的电力系统网络异常检测与分类方法视为方法 2。3 种方法的异常检测率如图 2 所示。

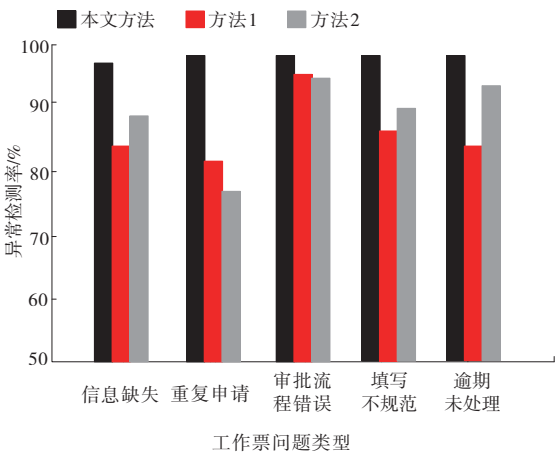


图 2 3 种方法的异常检测率

从图 2 中可以清晰地看到,基于随机森林算法在异常检测率上显著优于其他两种方法。该方法能够准确识别出工作票中的常见异常问题。相比之下,其他两种方法在检测某些特定类型的异常时可能表现出一定的局限性或不足。这一结果充分验证了本文方法的有效性和实用性。它不仅在理论上具有坚实的支撑,更在实际应用中展现出了卓越的性能。通过该算法的应用,可以更加高效地识别和处理工作票中的异常问题,为电力营销的现场管理和安全运营提供有力保障。

6 结语

本研究基于随机森林算法,对电力营销现场工作票中的问题进行了深入的问题分类与异常检测。该方法能够显著提高异常检测的准确性,及时揭示工作票中的异常情况,为电力企业的运维管理提供有力的技术支撑。但由于工作票数据的多样性和复杂性,尽管进行了大量的数据预处理和特征选择工作,仍有部分异常事件未能被模型准确捕捉。这提示在未来的研究中需要进一步优化特征工程,提高模型的泛化能力。未来将努力扩大数据集的规模和多样性,以涵盖更多类型的异常事件,进一步提升模型的实用性,为电力系统的安全稳定运行贡献更多的技术力量。

参考文献:

[1] 王楠,周鑫,周云浩,等. 基于改进 SVM 算法的电力工程异常数据检测方法设计[J]. 电子设计工程,2024,32(4):162-166.

[2] 段笑晨,王延. 基于机器学习的电力系统网络异常检测与分类方法[J]. 自动化与仪器仪表,2023(5):100-105.

[3] 宋世军,樊敏. 基于随机森林算法的大数据异常检测

- 模型设计[J]. 吉林大学学报(工学版),2023,53(9): 2659-2665.
- [4] PARLAK B, UYSAL A K. A novel filter feature selection method for text classification: extensive feature selector[J]. Journal of Information Science, 2023, 49(1):59-78.
- [5] OLATEJU O O, OKON S U, IGWENAGU U T I, et al. Combating the challenges of false positives in AI-driven anomaly detection systems and enhancing data security in the cloud[J]. Asian Journal of Research in Computer Science, 2024, 17(6):264-292.
- [6] ELIA G, GHIANI G, MANNI E, et al. A system for anomaly detection in reverse logistics: an application into an e-commerce company [J]. Measuring Business Excellence, 2024, 28(2):222-242.
- [7] VALENZUELA P L, MONTALVO-PEREZ A, ALEJO L B, et al. Are unilateral devices valid for power output determination in cycling? insights from the favero assioma power meter[J]. Int J Sports Physiol Perform, 2022, 17(3):484-488.
- [8] 王子馨, 胡俊杰, 刘宝柱. 基于长短期记忆网络的电力系统量测缺失数据恢复方法[J]. 电力建设, 2021, 42(5):1-8.
- [9] RAHMAN M M, MIM M A, CHAKRABORTY D, et al. Anomaly-based intrusion detection system in industrial IoT-healthcare environment network [J]. Journal of Engineering Research and Reports, 2024, 26(6):113-123.
- [10] 哈冠雄, 贾巧雯, 陈杭, 等. 无第三方服务器的基于数据流行度的加密去重方案[J]. 通信学报, 2022, 43(8):17-29.
- [11] AVANCINI D B, RODRIGUES J J P C, RABÊLO R A L, et al. A new IoT-based smart energy meter for smart grids [J]. International Journal of Energy Research, 2021, 45(1):189-202.
- [12] 张淑清. 基于哈希计算的大数据冗余消除算法设计[J]. 微型电脑应用, 2021, 37(12):68-70.
- [13] GANAGAVALLI K, SANTHI V. YOLO-based anomaly activity detection system for human behavior analysis and crime mitigation [J]. Signal, Image and Video Processing, 2024, 18(1):417-427.
- [14] HUANG Z, YANG S, ZOU Q, et al. A portable household detection system based on the combination of bidirectional LSTM and residual block for automatic arrhythmia detection[J]. Biomed Tech (Berl), 2024, 69(2):167-179.
- [15] 殷鹏程, 谭曼丽莎, 曹阳梅, 等. 基于迁移学习图像识别的桥梁监测数据异常检测方法[J]. 大连交通大学学报, 2024, 45(3):106-113.
- [16] PIEKARSKI M, JAWOREK - KORJAKOWSKA J, WAWRZYŃIAK A. Automatic analysis and anomaly detection system of transverse electron beam profile based on advanced and interpretable deep learning architectures[J]. Journal of Artificial Intelligence and Soft Computing Research, 2024, 14(2):139-156.
- [17] SAAD R M A, ANBAR M, MANICKAM S, et al. Neighbor discovery protocol anomaly-based detection system using neural network algorithm[J]. International Journal of Information Security, 2024, 23(3):1787-1803.
- [18] 范文明. 基于多源异构数据的高速列车状态智能检测系统[J]. 自动化技术与应用, 2024, 43(4):128-131.
- [19] ABUALGHANAM O, ALAZZAM H, ALHENAWI E, et al. Fusion-based anomaly detection system using modified isolation forest for internet of things [J]. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(1):131-145.
- [20] 何肖蒙, 王颖舒, 袁舒, 等. 基于 Wolf 的数字化变电站通信网异常流量检测系统[J]. 电子设计工程, 2024, 32(7):110-114.

收稿日期:2024-11-22