

杨肖珂,冯典,何林,等. 基于 KNN 算法的区域气象站电源故障预测模型研究应用[J]. 陕西气象, 2025(4): 69-73.

文章编号:1006-4354(2025)04-069-05

基于 KNN 算法的区域气象站电源故障预测模型研究应用

杨肖珂¹,冯典²,何林³,杨少康¹

(1. 渭南市气象局,陕西渭南 714000;2. 陕西省气象台,西安 710014;

3. 陕西省气象信息中心,西安 710014)

摘要:电源故障已成为影响区域气象站数据及时性和完整性的主要原因。为降低维护成本并提高故障预测的准确率,提出基于 K-最近邻(K-Nearest Neighbor, KNN)算法的电源故障预测模型。利用渭南市 2019 年 11 月至 2022 年 5 月 85 个因电源系统故障导致的区域气象站报文缺测记录,提取每例故障发生前 10 d 回传电压数据标记为“故障”样本,再随机选取该时段内 335 条无电源异常的连续 10 d 的逐小时电压数据标记为“正常”样本,共计 420 条数据作为训练与测试样本。计算每条数据的统计学特征并归一化处理后再优选特征子集,最后训练并评估 KNN 分类器。试验结果表明,所构建的 KNN 模型准确率(Accuracy)为 0.95、精确率(Precision)为 0.95、召回率(Recall)为 0.81、F1-score 为 0.88、AUC 值为 0.90。2022 年 9 月至 12 月的实际应用结果显示,该模型成功地提前标记了所有故障实例(漏报率为 0),但在连续低温寡照的天气条件下,该模型显示出较高的误报率。该方法对于提高电源故障预判能力具有一定的应用价值,也为后续进一步结合天气要素、站点环境等多源信息建立精细化模型提供了参考。

关键词:KNN 算法;区域气象站;电源故障;预测模型;机器学习

中图分类号:P415.12

文献标识码:A

区域气象站作为全天候、全自动运行的气象观测设备,在中小尺度天气系统监测、精细化预报服务等方面发挥了重要作用。传感器故障、电源故障和通信传输故障是区域气象站的三大主要故障类型,其中电源故障已成为影响区域气象站数据及时性和完整性的主要原因^[1-2]。目前大多数区域气象站采用太阳能供电,铅酸免维护电池通过电源控制器实现整站供电及蓄电池充放电控制^[3-4]。然而,受天气状况(低温、阴雨寡照等)、站点周边环境(树木、建筑物遮挡)以及电池老化、线路故障等多重因素影响,电源故障频发。为了减少因电源故障造成的数据缺测,提高现场维护的有效性,基于历史监测数据开展电源故障预判具

有实际意义。传统方法多依赖简单阈值报警或基于电池剩余容量与放电电流关系的估算^[5-6],但這些方法难以适应不同季节、气象要素变化和站点差异等复杂情况。随着数据科学和机器学习技术的发展,利用历史数据挖掘故障模式并建立预测模型成为一种可行的解决方案。采用 K-最近邻(K-Nearest Neighbor, KNN)算法构建故障预测模型,通过对区域气象站回传的电压数据进行分析,可以有效预警电源可能出现的故障。KNN 算法因其简单易用且能处理非线性问题的特点,特别适合用于此类应用场景。通过这种方式,技术人员能够提前识别潜在问题,从而优化电源故障处理流程,确保区域气象站的稳定运行。

收稿日期:2023-06-12

作者简介:杨肖珂(1982—),男,汉族,陕西富平人,本科,工程师,主要从事气象信息技术保障。

基金项目:渭南市气象局科研项目(2022wk-3)

1 数据与方法

1.1 数据来源与样本构建

渭南市区域气象站 2019 年 11 月至 2022 年 5 月期间共出现 85 例因电源系统故障导致的数据缺失。为了分析故障的发生规律,从每例故障站点中提取故障发生前连续 10 d 的逐小时回传电压数据,标记为“故障”样本。同时,为了构建与“故障”样本相对比的“正常”样本,在相同时间段内随机选取 335 条无电源异常的站点数据(具备完整的逐小时观测记录,并且没有出现明显的电压缺失),同样提取连续 10 d 的逐小时电压数据,标记为“正常”样本。

为了确保“正常”样本与“故障”样本在时间分布和其他特征上的匹配性,在选取过程中尽量保持两组样本的时间跨度一致,并考虑了不同季节、天气条件下的代表性。选择 10 d 的时间窗口是基于经验观察和初步数据分析,旨在捕捉电源系统在故障发生前可能出现的异常变化趋势。

1.2 数据预处理与特征提取

1.2.1 数据清洗与异常值处理 在获取前 10 d 逐小时电压序列后,首先对其中的缺失值和明显不合理值(如极端尖峰或负值)进行清洗。对于缺失值,如果缺失量较大(例如超过连续 2 d 的数据缺失),则直接剔除此条记录。对于少量缺失值,采用线性插值或其他合适的插值方法进行填补。此外,通过设定合理的阈值(如电压值应在正常工作范围内),剔除超出该范围的异常值。

1.2.2 统计学特征提取 针对时间序列数据的特征提取,统计学方法因其简单性和物理意义明确,在故障识别方面常能获得较好效果。综合参考文献^[7]和前期试验结果,选取以下特征进行提取:近 10 d 电压值的最小值(Min)、近 10 d 电压值的最大值(Max)、近 10 d 电压值的平均值(Mean)、近 10 d 电压值的方差(Var)、近 10 d 电压值的标准差(Std)、近 10 d 电压值的变化趋势(Grad)、线性回归计算的电压斜率(Slope)、线性回归拟合方程的截距(Intercept)、电压序列中明显的峰值数量(Peaks,基于一定阈值或极值检测算法判定)。这些特征能够全面反映电压序列的变化趋势和波动情况,有助于提高故障识别的准

确性。

1.2.3 归一化处理 由于上述特征值的量纲不同,为了便于 KNN 算法的距离计算,需要将特征统一缩放到相同的数值区间。采用 MinMaxScaler(Min-Max 归一化公式)将每个特征转换为 $[0, 1]$ 区间。具体公式如下:

$$x' = (x - x_{\min}) / (x_{\max} - x_{\min}), \quad (1)$$

其中, x 为原始特征值, $x_{\min}$ 与 $x_{\max}$ 分别为该特征的最小值与最大值, x' 为转换后的特征值。通过这种方式,确保了不同特征之间具有可比性,从而提高了 KNN 算法的性能。

1.3 KNN 算法

1.3.1 原理 KNN(K-Nearest Neighbors)算法是一种基于实例的监督学习方法,其基本原理是通过计算新样本与训练集中所有样本的距离,找到距离最近的 K 个邻居,然后根据这些邻居的多数类别来判定新样本的类别。KNN 分类算法操作简单,对小规模数据具有较好的分类效果,常作为数据分类的初步尝试方法^[8-10]。

1.3.2 距离度量与权重 本研究选用欧氏距离作为相似度度量方式。欧氏距离是一种常用的衡量两个样本之间相似度的方法,它通过计算两个样本在各个特征维度上的差值平方和的平方根来得到。为了简化模型,采用统一权重的方式,即所有邻居样本对分类投票的权重相同。在实际应用中发现,两种权重方式的效果差异不大,因此选择统一权重来简化模型。

1.3.3 特征选择与交叉验证 为进一步优化特征子集并避免过拟合,采用递归特征消除结合交叉验证方法对 9 个特征进行递归特征消除。具体做法为:在五折交叉验证的基础上,每次训练 KNN 模型后,根据特征的重要性或预测效果剔除贡献最小的特征,直至找到表现最优的特征子集。实验结果显示,在保留 6~7 个特征时,模型评估指标(如准确率、F1-score 等)达到最优或接近最优。

1.4 电压特征

根据区域气象站电源工作原理,电源系统正常工作时,蓄电池的空载电压约为 13.5 V,在带负载的情况下工作电压一般不会低于 12 V。在

晴朗天气下,太阳能板(也称为光伏板)直接向设备供电,并补充无日照期间放电所消耗的电量,使蓄电池电压上升;而在夜间或阴雨天无有效日照时,蓄电池放电供设备用电,导致蓄电池电压因放电而缓慢下降。引起区域气象站电源系统故障的原因有很多,可以通过分析回传电压值的变化来加以识别。以下是几种常见的故障情况及其对应的电压变化特征。

(1) 蓄电池性能下降

蓄电池性能下降时,阴极板可能发生不可逆硫酸铅化,有效容量变小。实际使用中表现为蓄电池充电迅速达到满电状态,但在放电过程中电压值迅速下降(图 1a)。不可逆硫酸铅化是由于电池长期处于部分充电状态或深度放电状态,导致硫酸铅晶体沉积在电极表面,降低了电池的有效容量。

(2) 充电不足

当区域气象站周围有树木、建筑物等遮挡物,或者连续低温寡照天气会影响太阳能电池板的充

电效率,导致蓄电池充电不足。在低温状态下,蓄电池的容量和性能都会显著降低,如果再叠加连续阴雨雪天气,蓄电池长时间处于亏电状态,加速了电池老化(图 1b)。

(3) 太阳能电池板或电源控制器故障

有时太阳能电池板或电源控制器出现故障,如充电效率下降或无法充电,最终导致电池亏电宕机(图 1c)。此外,接线头松动或锈蚀也会增加电阻,减少电流强度,从而影响充电效果。

(4) 控制器低压切断保护功能

电源控制器具有低压切断保护功能,避免深度放电。当蓄电池电压降至 11.0 V 时,控制器会自动切断负载输出。在实际数据分析中发现多个站点的回传电压值低于 11.0 V(图 1d),这种情况可能是由于设备保障人员为了保证站点持续运行,避免被控制器的深放电保护切断,人为干预将负载直接连接到蓄电池上。这种操作虽然暂时维持了站点运行,但也增加了电池过度放电的风险,进一步损害电池寿命。

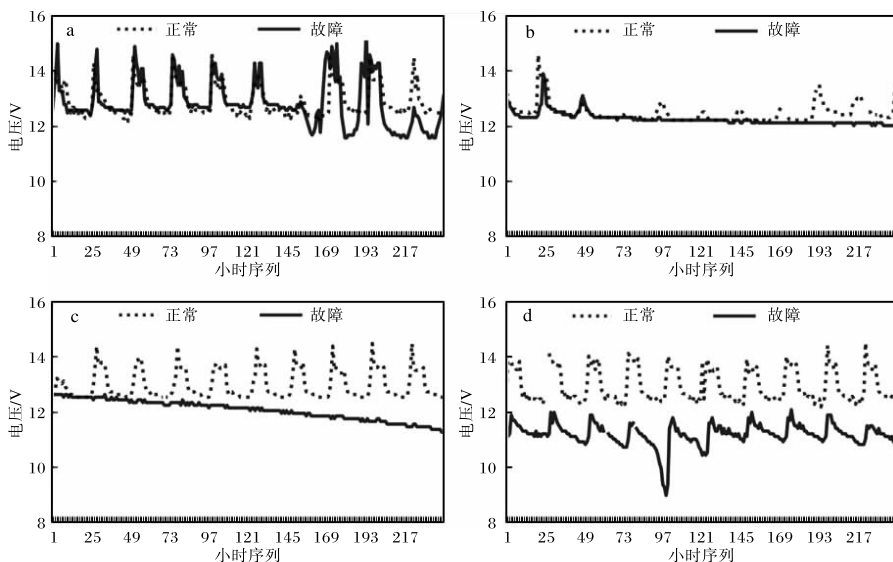


图 1 几种较典型回传电压变化情况

(a 蓄电池性能下降;b 充电不足;c 硬件故障;d 其他)

2 建模及检验

2.1 使用 KNN 算法构建预测模型

Scikit-learn 是一个基于 Python 语言的机器学习工具包,广泛应用于统计分析和机器学习建模等数据科学领域。选用 KNN 算法构建分类器

模型,并使用 Scikit-learn 进行模型的训练和预测。为了优化特征子集,采用 Scikit-learn 中的 RFECV(递归特征消除结合交叉验证)函数进行特征选择。K 值的大小直接影响分类边界的平滑度和模型对局部异常点的敏感度,为了确定合适

的 K 值,在不同的随机数种子下多次划分训练集和测试集,并通过五折交叉验证的方法评估模型性能。图 2 展示了不同 K 值对应的预测模型准确率,横轴表示 K 值,取值范围为 1~10,纵坐标表示准确率。实验结果表明,当 K 值为 3 时,预测模型的准确率达到最高,约为 0.94。

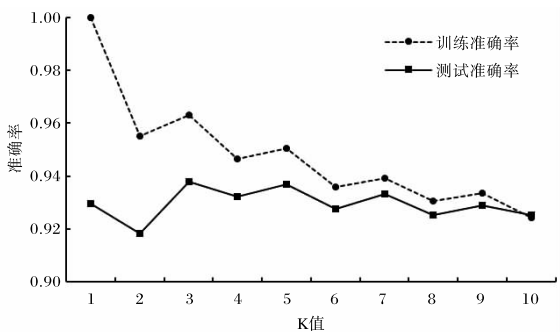


图 2 不同 K 值下预测模型的准确率

2.2 模型评价

为了全面评估模型的预测效果,使用了多个评价指标,包括准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1-score 和 AUC(表 1)。

表 1 KNN 模型评价指标得分

Accuracy 值	Precision 值	Recall 值	F1-score 值	AUC 值
0.95	0.95	0.81	0.88	0.90

2.3 实际应用案例

将训练好的 KNN 模型应用于 2022 年 9 月至 12 月的业务监测系统中,每日提取站点回传电压并统计特征值,输入模型后生成“正常”或“故障”风险标记。在此期间,共有 9 个站点先后出现了因电源系统故障导致的数据缺测,模型在故障发生前均给出了预警标记,实现了零漏报。然而,在连续阴雨寡照或低温天气下,部分站点的电压曲线显著下降,模型判定为“故障”风险。尽管后期天气转好、电池得到充分充电后并未真正发生宕机,但在这种情况下仍导致了较高的误报率。峰值时,警告状态数量可达总站点数的 12% 至 15%。

总体来看,“零漏报”对于业务的安全性至关重要,但在特殊天气条件下(如连续阴雨或低温),

准确率是指分类模型中所有判断正确的结果占总观测值的比重,本模型的准确率为 0.95,表明在所有样本中,95% 的样本被正确分类。精确率是指在模型预测为正类的所有结果中,实际为正类的比例,本模型的精确率为 0.95,表明在所有预测为正类的结果中,95% 是正确的。召回率是指在实际为正类的样本中,被模型正确预测为正类的比例,衡量的是模型的查全率,本模型的召回率为 0.81,表明在所有实际为正类的样本中,81% 的样本被正确识别。F1-score 是精确率和召回率的调和平均值,综合考虑了模型的查准率和查全率,本模型的 F1-score 为 0.88,表明模型具有较高的可靠性。AUC 是指 ROC 曲线下的面积,介于 0.5 到 1.0 之间,AUC 越接近 1.0,表示模型的分类性能越好。本模型的 AUC 为 0.90,表明模型具有很高的分类准确性。综合以上评价指标来看,预测模型在各项指标上均表现良好,尤其是在准确率、F1-score 和 AUC 等关键指标上取得了较好的平衡,具有较高的预测效果和可靠性。

模型的误报率较高,这表明需要进一步优化模型以降低误报率。

3 结论与讨论

(1)利用渭南市 2019 年 11 月至 2022 年 5 月期间区域气象站故障记录与回传电压数据,构建基于 KNN 算法的电源故障预测模型,通过提取近 10 d 电压序列的统计学特征,并进行归一化处理后,采用递归特征消除结合交叉验证方法对模型进行了优化。实测结果表明,基于 KNN 算法的电源故障预测模型具有较好的预测精度。该模型在准确率(Accuracy)、精确率(Precision)、F1-score 及 AUC 等指标上均表现良好,尤其在在实际应用中能有效识别大多数即将发生的电源故障。

(2)该模型仅依赖回传电压这一单一数据信息,即可完成对区域气象站电源故障风险的初步

判定。这种方法特别适用于缺少其它辅助监测参数的站点,具有快速部署、可推广性强的特点,并为更大范围、更长时段的推广应用奠定了基础。

(3)在恶劣气象条件下(如连续阴雨或低温),模型会产生一定的误报。未来可通过扩充样本规模、引入多源信息(如气象预报、站点环境条件)与高阶特征提取方法,提高模型的准确性和稳健性,从而更好地支持区域气象站的故障预测和预防性维护。

参考文献:

- [1] 吴玉芳,丁媛. 区域气象站电源系统故障的判断与排除[J]. 现代农业科技,2012(11):228.
- [2] 董彭康. 区域气象站电源系统故障的判断与排除[J]. 资源环境,2012(11):155.
- [3] 江苏省无线电科学研究所有限公司. DZZ4 型自动气象站用户手册[Z]. 2014.
- [4] 中国华云技术开发公司. DZZ5 型自动气象站用户指南及系统说明书[Z]. 2012.
- [5] 陈澄. 区域自动气象站电源系统报警阈值研究[J]. 气象水文海洋仪器,2020(6):11-18.
- [6] 季国栋,王文理. 蓄电池剩余容量检测系统设计[J]. 微型机与应用,2015(3):42-44.
- [7] 林珠. 数据挖掘中适用于分类的时序数据特征提取方法[J]. 计算机系统应用,2012(10):224-228.
- [8] HARTIGAN J A, WONG M A. Algorithm AS 136: A k-means clustering algorithm[J]. Journal of the Royal Statistical Society Series C (Applied Statistics), 1979, 28(1): 100-108.
- [9] 顾文浩. 基于改进 KNN 算法的股票风险预测[J]. 现代商业,2022(18):157-160.
- [10] 窦小凡. KNN 算法综述[J]. 通讯世界,2018(10): 273-274.