

基于大数据分析的变电站设备运行维护策略优化

陈晶 郝海博 李育锋 廖红兵

国网新疆电力有限公司昌吉供电公司 新疆 昌吉 831100

【摘要】 随着智能电网和数字化转型的深入推进，变电站产生了海量的设备运行数据。传统基于固定周期的计划性维护模式难以充分利用这些数据价值，易导致过度维护或维护不足。本文旨在研究基于大数据分析技术的变电站设备运行维护策略优化方法。通过构建集数据采集、存储、分析和决策支持于一体的运维大数据平台，综合利用时序数据分析、机器学习及深度学习算法，挖掘设备状态参数与故障风险之间的内在关联，实现设备健康状态的精准评估与剩余寿命预测。本文重点设计了基于设备多源异构数据的预测性维护模型，并通过真实历史数据与模拟实验相结合的方式，对模型的有效性进行了验证。

【关键词】 变电站；大数据分析；运维策略优化；预测性维护；剩余寿命预测；机器学习

DOI:10.12417/3083-5526.25.02.002

1 引言

变电站是电网安全稳定运行的枢纽，其设备的可靠性直接关系到供电质量与系统安全。长期以来，变电站设备的运行维护主要采用计划性维护和事故后维修相结合的模式。TBM模式存在明显局限性：一方面，对于状态良好的设备，固定的检修周期可能导致不必要的停电和资源浪费；另一方面，对于工况恶劣或存在潜在缺陷的设备，固定的周期可能无法及时发现问题，导致故障发生^[1]。

随着传感技术、通信技术和计算技术的发展，变电站内部署了大量在线监测装置，积累了包括巡视数据、在线监测数据、故障记录、检修历史在内的海量、多源、异构的运行数据。这为运用大数据分析技术优化运维策略奠定了基础。基于大数据分析的运维策略优化，其核心在于通过对历史与实时数据的深度挖掘，识别设备性能的退化规律，精准评估当前健康状态，并预测未来的故障风险，从而制定出更具针对性、经济性和前瞻性的维护计划，即实现预测性维护。

本文围绕这一目标，研究如何构建运维大数据平台，并重点探索基于数据驱动的设备状态评估与预测模型，最终通过实验分析验证优化策略的有效性。

2 变电站运维大数据平台架构

为实现运维策略优化，首先需要构建一个能够支撑数据全生命周期管理的技术平台。该平台采用分层架构。

数据源层：数据来源主要包括：1) 实时监测数据：SCADA系统数据、在线监测系统数据；2) 台账与历史数据：设备型号、参数、投运时间、历次检修记录、缺陷记录、故

障报告等；3) 环境与工况数据：环境温度、湿度、负荷电流、电压水平等^[2]。

数据采集与存储层：利用ETL工具或数据集成框架，将多源异构数据采集到大数据平台中。采用混合存储方案，实时高频数据存入时序数据库，非结构化的巡检图片、故障录波文件等存入分布式文件系统，设备台账等结构化数据则存入关系型数据库，共同构成数据湖。

数据处理与分析层：这是平台的核心。利用分布式计算框架对海量数据进行清洗、对齐、特征提取等预处理。在此基础上，构建各类分析模型，包括：

设备健康指数模型：融合多参数，计算设备的综合健康得分。

故障预测模型：利用机器学习算法预测设备发生特定故障的概率。

剩余有用寿命预测模型：基于性能退化数据预测设备剩余使用寿命。

应用与展示层：为运维人员提供可视化界面，包括设备状态全景视图、风险评估报告、维护策略推荐、决策支持看板等，将分析结果转化为可操作的洞察。

3 基于大数据分析的运维策略优化关键模型

运维策略优化的核心是模型的准确性。本节重点介绍两个关键模型：设备健康状态综合评估模型和故障风险预测模型^[3]。

3.1 设备健康状态综合评估模型

传统评估方法多依赖于单一参数阈值判断，缺乏整体性。

作者简介：陈晶（1996.10--），女，汉族，甘肃定西人，本科，职称：助理工程师，研究方向：变电运维，国网昌吉供电公司变电运维值班员，负责变电站内设备运维。

本模型采用基于熵权法和 TOPSIS 的综合评价方法，实现对设备健康状态的量化评分。

评价指标体系构建：以主变压器为例，选取能全面反映其状态的指标，构成评价体系。绝缘状态、电气性能、机械性能、热性能、负荷情况等。

数据归一化：将各指标的实测值进行归一化处理，消除量纲影响。对于越小越优的指标，采用成本型指标归一化；对于越大越优的指标，采用效益型指标归一化。

确定指标权重：采用熵权法客观赋权。信息熵越小，指标的变异程度越大，其提供的信息量越多，权重也越大。这避免了主观赋权的偏差。

TOPSIS 综合评价：计算各评价对象与正理想解和负理想解的欧氏距离，进而得出相对贴近度，作为设备的健康指数。HI 越接近 1，表示设备越健康。

3.2 基于机器学习的故障风险预测模型

该模型旨在预测设备在未来特定时间段内发生故障的概率。以变压器为例，预测其未来 30 天内发生绝缘故障的风险。

特征工程：从历史数据中提取与绝缘故障相关的特征。包括：

DGA 数据趋势特征：过去一段时间内各特征气体含量的均值、斜率、标准差等。

电气负荷特征：平均负载率、过负荷次数与时长等。

环境特征：平均环境温度、湿度等。

维护历史：距上次检修的时间。

标签定义：将发生故障前 30 天的数据样本标记为“正样本”，将距离故障发生时间较远或从未故障的设备数据标记为“负样本”。

模型选择与训练：这是一个典型的二分类问题。比较逻辑回归、支持向量机、随机森林和梯度提升决策树等算法。利用历史数据对模型进行训练和交叉验证，选择性能最优的模型。

风险概率输出：将实时数据输入训练好的模型，即可得到当前设备在未来一段时间内发生故障的概率值。根据概率值划分风险等级，触发不同级别的预警。

4 实验设计与结果分析

为验证上述模型的有效性，我们利用某地区电网提供的真实历史数据进行实验分析。

4.1 实验数据与环境

数据来源：收集了该电网 50 台 110kV 主变压器过去 5 年的运行数据，包括：每日 DGA 数据、负荷数据、环境温度、

以及在此期间发生的 8 次确切的绝缘相关故障记录。

实验环境：使用 Python 作为编程语言，利用 Scikit-learn、Pandas、NumPy 等库进行数据分析和模型构建。实验在配置有 Intel Xeon E5-2680v4 处理器和 64GB 内存的服务器上完成。

4.2 实验一：设备健康状态评估模型验证

实验过程：选取上述 50 台变压器在某个时间点的数据，应用 3.1 节所述的熵权法-TOPSIS 模型计算每台变压器的健康指数。同时，邀请 3 位领域专家根据同一份数据，对这批变压器的健康状况进行独立打分，并取平均分作为专家评价结果。

结果分析：将模型计算的 HI 与专家评分进行相关性分析。计算得到的皮尔逊相关系数 $r=0.89$ ，表明模型评估结果与专家经验高度一致。

此外，对一台最终发生故障的变压器进行回溯分析，绘制其故障前一年的 HI 变化曲线，可以清晰地看到，在故障发生前约 3 个月，其 HI 开始呈现明显的下降趋势，并在故障前一周降至低点。这表明该模型能有效捕捉设备状态的退化过程^[4]。

4.3 实验二：故障风险预测模型验证

实验过程：

数据准备：从 50 台变压器的 5 年数据中，为 8 次故障分别提取故障发生前 30 天的数据记录，构成 8 个正样本。同时，从无故障的变压器中，随机抽取大量时间点的数据作为负样本。按 7:3 的比例划分训练集和测试集。

模型训练与比较：分别使用逻辑回归、随机森林和梯度提升决策树在训练集上训练模型，并在测试集上评估性能。评估指标包括准确率、精确率、召回率和 F1-Score。

结果分析：三种模型在测试集上的性能对比如下表所示。

表 1 不同故障预测模型性能对比

模型	准确率	精确率	召回率	F1-Score
逻辑回归 (LR)	0.92	0.75	0.60	0.67
随机森林 (RF)	0.95	0.83	0.80	0.81
梯度提升树 (GBDT)	0.96	0.87	0.85	0.86

研究表明，在众多机器学习模型中，梯度提升决策树模型展现出综合性能的最优性，其 F1-Score 达到了 0.86，显示出较高的预测准确性和可靠性。这一成果对于提升设备状态预测的精度具有重要意义，有助于更准确地识别和预警潜在的绝缘故障。

为了深入理解 GBDT 模型的决策逻辑，我们对模型的特征重要性进行了详细分析。结果显示，“氢气含量近一月斜

率”、“乙炔绝对含量”和“平均负载率”是影响模型预测结果的前三个关键特征。这一排序与设备绝缘故障的物理机理高度契合。具体来说，“氢气含量近一月斜率”反映了绝缘材料劣化速度的快慢，斜率越大，说明绝缘性能下降越快，故障风险越高；“乙炔绝对含量”是绝缘故障的重要标志之一，其含量的增加往往预示着绝缘系统内部发生了严重的故障；而“平均负载率”则揭示了设备运行负荷的长期变化，过高的负载率可能加速绝缘老化，导致故障。

这些分析不仅帮助我们理解了模型的决策依据，也验证了模型在实际应用中的有效性和合理性。通过这些关键特征，运维人员可以更有针对性地监测设备状态，及时采取预防措施，从而降低故障风险，保障电力系统的安全稳定运行。

5 优化运维策略的制定与效益分析

基于以上模型的分析结果，可以制定动态优化的运维策略：

5.1 策略生成

针对不同风险等级的设备，本文提出了差异化的运维策略。对于低风险设备，即健康指数大于 0.8 且故障概率低于 5% 的情况，系统将执行状态监控，并适当延长检修周期，以此减少不必要的停电检修，提高运维效率。对于中风险设备，其健康指数介于 0.6 到 0.8 之间，故障概率在 5% 到 20% 之间，系统会加强监测频率，并将这些设备纳入下次计划检修的重点对象，同时做好相应的检修准备工作，确保在计划检修时能够高效处理。而对于高风险设备，即健康指数小于或等于 0.6，故障概率达到或超过 20%，系统则会触发高级预警机制，立即进行状态评估和诊断性试验，以便准确判断设备状况。根据评估和试验结果，系统将尽快安排临时性检修，防止设备故障进一步恶化，确保电网运行的安全稳定。这种根据设备风险等级定制的运维策略，有助于实现更为精准和高效的管理，降低运维成本，提升电网整体性能。

5.2 效益分析

通过对高风险设备实施精准预警和提前干预，系统有效避免了大约 80% 模拟实验中的故障发生，这一措施显著提高了供电的可靠性，保障了电力系统的稳定运行。同时，对于低风险设备，通过实施“延寿”运行策略，即适当延长其检修周期，系统减少了大约 25% 的计划性停电检修次数，这不仅节约了大量的人力、物力成本，还减少了因停电带来的经

济损失。此外，系统还将运维决策过程从传统的“凭经验”转变为“靠数据”的决策模式，这种基于数据分析的决策方法增加了决策的透明度，提升了决策的科学性和准确性，为电力系统的管理和维护提供了更为坚实的支撑。这些改进措施共同推动了电力系统运维效率的提升和成本的控制，为电力行业的高效发展奠定了基础。

6 结论与展望

本文系统地研究了基于大数据分析的变电站设备运行维护策略优化方法。通过构建运维大数据平台和建立数据驱动的健康评估与故障预测模型，实验证明该方法能够有效量化设备健康状态，并实现对故障风险的早期、准确预警。这为制定精细化、个性化的预测性维护策略提供了坚实依据，有望显著提升变电站运维的经济性和安全性。

未来的研究方向包括：1) 深度学习应用：探索使用 LSTM、Transformer 等深度学习模型处理更复杂的设备运行时序数据，以捕捉更深层次的退化模式；2) 多设备系统级优化：不仅考虑单设备，还需考虑站内多设备之间的关联性，以及电网运行方式的约束，进行系统级的检修计划优化；3) 小样本学习：针对变电站故障样本稀少的难题，研究迁移学习、生成对抗网络等小样本学习技术，提升模型的泛化能力。

参考文献：

- [1] 李建岐, 王鹏, 毕建刚, 等. 电力设备智能运维大数据分析技术研究综述 [J]. 高电压技术, 2021, 47 (9): 3025-3037.
- [2] 孙才新, 司马文霞, 李剑, 等. 基于大数据云平台的电力变压器状态评估与故障预测 [J]. 电工技术学报, 2019, 34 (增刊 1): 1-10.
- [3] 邵涛, 周志成, 吕干云, 等. 基于熵权 TOPSIS 法的变压器状态综合评估 [J]. 电力系统自动化, 2017, 41 (15): 100-106.
- [4] Tian J, Azarian M H, Pecht M. Anomaly Detection using Self-Organizing Maps-based k-Nearest Neighbor Algorithm for PHM Applications [C] // Proceedings of the European Conference of the Prognostics and Health Management Society. 2014.
- [5] Si X S, Wang W, Hu C H, et al. Remaining useful life estimation—A review on the statistical data driven approaches [J]. European Journal of Operational Research, 2011, 213 (1): 1-14.