

# 智能电网变电运维中的大数据分析技术应用研究

吕晓帆 窦艳艳 帕里扎提·帕尔屯 高 婷

国网新疆电力有限公司昌吉供电公司 新疆 昌吉 831100

**【摘要】：**本文探究了大数据分析技术在智能电网变电运维中的应用，构建了一个全面的技术框架，并聚焦于设备状态评估、故障预警、运维策略优化和运行方式智能决策四个关键场景。针对变压器，研究设计了基于时序数据挖掘和机器学习的故障预测实验，并使用真实数据进行了验证。实验证明，该技术方案能有效地挖掘数据价值，提前识别设备异常，为提升变电运维的安全性和效率提供了技术支持。

**【关键词】：**智能电网；变电运维；大数据分析；故障预测；状态评估；深度学习

DOI:10.12417/2811-0536.26.01.056

## 1 引言

智能电网的核心特征在于电网的智能化、信息化和自动化。作为智能电网的核心环节，变电站内部署了海量的智能传感器、监测终端和控制系统，每时每刻都在产生包括SCADA数据、在线监测数据、视频图像、巡检记录、故障录波等在内的巨量数据。这些数据蕴含着反映设备健康状况、电网运行状态和潜在风险的宝贵信息。然而，传统变电运维模式主要依赖人工经验分析有限的监测参数和阈值告警，难以应对数据爆炸式增长带来的挑战，存在信息孤岛、分析浅层、响应滞后等问题<sup>[1]</sup>。大数据分析技术，以其对海量数据的存储、管理和深度挖掘能力，为破解上述困境提供了全新的解决方案。它将变电运维从“被动响应”向“主动预警”、从“计划检修”向“状态检修”和“预测性维护”转变成为可能。通过应用数据挖掘、机器学习和深度学习等先进算法，可以从纷繁复杂的数据中提取有效特征、发现隐性规律、构建预测模型，最终实现设备状态的精准画像、故障的早期诊断、运维资源的优化配置以及运行方式的自适应调整。

本文旨在系统性地研究大数据分析技术在智能电网变电运维中的整体应用框架、关键技术与典型应用场景，并通过具体的实验分析验证其在提升运维智能化水平方面的有效性与价值。

## 2 变电运维大数据分析技术框架

一个完整的变电运维大数据分析体系通常包含以下五个层次。

(1) 数据感知与采集层：这是数据来源的根基。通过部署于站内的各类传感器、监测装置、保护控制设备以及机器人/无人机巡检系统，全面采集设备状

态、电气量、环境参数等结构化、半结构化和非结构化数据<sup>[2]</sup>。

(2) 数据集成与存储层：负责将多源异构数据汇聚、清洗、整合。利用数据总线、消息队列等技术实现数据的实时/准实时接入。采用混合存储架构，如时序数据库用于存储高频监测数据，关系型数据库存储设备台账和配置信息，分布式文件系统或对象存储用于存放图像、视频等非结构化数据，共同构成变电运维数据湖<sup>[3]</sup>。

(3) 数据处理与分析层：这是整个框架的“大脑”。利用分布式计算引擎对海量数据进行高效批处理和流处理。在此基础上，集成各类数据分析算法：①数据预处理：数据清洗、缺失值填补、数据归一化等。②统计分析：描述性统计、相关性分析、趋势分析等。

(4) 机器学习/深度学习：用于分类、回归、聚类、异常检测等。

(5) 智能应用层：智能应用层作为整个系统的高级阶段，负责将深入分析的结果转化为可操作的业务应用。该层主要包括设备状态全景评估与预警，通过综合分析各类数据，实现对设备状态的全面监测和实时预警，确保设备的健康运行；故障预测与健康管理，运用先进的数据分析和模型预测技术，对潜在故障进行预测，并提供健康管理方案；运维策略优化与决策支持，根据分析结果优化运维计划，为管理人员提供科学的决策依据；以及电网运行方式优化建议，提出改进电网运行的策略，提高电网效率。而可视化与人机交互层则承担着将复杂的数据分析结果转化为直观、友好的信息展示的任务，通过Web端、移动APP等平台，以驾驶舱式的总览、趋势曲线的动态展现、

作者简介：吕晓帆，女（1998.03--），汉族，新疆昌吉市人，学历：本科，职称：助理工程师，研究方向：输变电运维，从事35-220千伏变电站变电运维专业5年。

三维可视化的空间布局、以及详细的报表等形式，将分析结果直观地呈现给运维管理人员。这种交互方式不仅提高了信息的可读性和易理解性，而且极大地提升了管理人员的工作效率，使其能够基于这些信息进行更为精准和高效的决策，从而保障整个电力系统的稳定和高效运行<sup>[4]</sup>。

### 3 大数据分析在变电运维中的核心应用场景

#### 3.1 设备状态精准评估与异常检测

传统阈值告警易产生误报和漏报。大数据分析通过多参数融合和模式识别，实现更精准的状态评估。

技术方法：采用主成分分析、隔离森林、自编码器等无监督或半监督学习算法，学习设备正常运行时的数据模式。当新输入的数据与学到的正常模式产生显著偏差时，即判定为异常。这种方法能够发现未知类型的故障和缓慢发展的隐性缺陷。

应用实例：对变压器 DGA 数据，结合负荷、油温等多维数据，利用 PCA 降维后，通过计算样本到正常数据分布中心的马氏距离来综合判断其状态，比单一气体阈值更可靠。

#### 3.2 故障预测与健康管理

这是大数据分析价值最突出的领域，旨在故障发生前进行预测。

技术方法：①基于传统机器学习：对历史故障数据和正常数据进行分析，提取特征，使用支持向量机、随机森林、梯度提升树等算法构建分类或回归模型，预测故障概率或剩余使用寿命。②基于深度学习：对于时序数据，利用长短期记忆网络或时序卷积网络能够更好地捕捉长期依赖关系和复杂动态特性。对于红外、可见光等图像数据，使用卷积神经网络进行缺陷自动识别。

应用实例：利用 LSTM 网络学习断路器分合闸线圈电流波形的时间序列特征，预测其机械故障的发生时间。

#### 3.3 运维策略优化与决策支持

基于预测结果，优化检修计划和资源分配。

技术方法：结合设备状态预测结果、电网运行方式、物资库存、人员排班等约束条件，建立优化模型，生成成本最低、可靠性最高的维护计划。同时，利用关联规则挖掘分析历史缺陷记录，发现不同设备缺陷之间的关联性，为综合性检修提供依据。

应用实例：根据多台变压器的健康状态和故障风险预测结果，综合考虑停电影响和检修成本，动态生

成未来一个季度或一年的最优检修序列。

#### 3.4 运行方式智能决策

通过对电网实时运行数据的分析，辅助运行人员做出最佳决策。

技术方法：利用数据驱动的方法或结合物理模型与数据的方法，进行潮流计算、静态安全分析、电压稳定评估等，并给出调整建议。

应用实例：基于历史数据和实时数据，训练神经网络模型快速评估不同运行方式下的系统稳定性，为调度员提供操作建议。

### 4 实验分析：基于大数据分析的变压器故障预测

为验证大数据分析技术在变电运维中的实际效果，以油浸式变压器为例，设计并完成了一个故障预测实验。

#### 4.1 实验目标

基于变压器历史的 DGA 数据序列，预测其在未来一段时间内发生绝缘故障的风险。

#### 4.2 数据准备与预处理

数据来源：实验数据来源于某电网公司提供的真实历史数据集，包含 150 台 110kV 及以上电压等级变压器长达 8 年的运行记录。其中，有 12 台变压器在此期间发生过经确认的绝缘故障，其余为正常变压器。

特征工程：①基础特征：选取  $H_2$ ,  $CH_4$ ,  $C_2H_6$ ,  $C_2H_4$ ,  $C_2H_2$  这 5 种关键特征气体的含量作为基础特征。②衍生特征：为捕捉气体变化趋势，计算了每个气体在过去 30 天、90 天内的平均含量、变化斜率、标准差等统计量。同时，计算了国际通用的故障诊断特征比值，如  $CH_4/H_2$ ,  $C_2H_2/C_2H_4$ ,  $C_2H_4/C_2H_6$  等。③标签定义：对于发生故障的变压器，将其故障发生前 30 天内的所有数据样本标记为“1”。对于正常变压器，随机抽取其正常运行期间的数据样本标记为“0”。最终构建了一个包含约 15,000 个样本的数据集。④数据预处理：对数据进行标准化处理，并采用 SMOTE 算法对正样本进行过采样，以解决样本类别不均衡问题。

#### 4.3 模型构建与对比

对比了三种典型的机器学习模型和一种深度学习模型：①逻辑回归：作为基线模型。②随机森林：能处理非线性关系，对特征重要性有较好解释。③梯度提升决策树：通常具有更高的预测精度。④长短期记忆网络：专门用于处理时序数据，有望捕捉气体含量的长期依赖关系。模型输入为过去 60 天每天的气体含量序列。

将数据集按 7:3 的比例随机划分为训练集和测试集。所有模型在训练集上训练，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 分数。

#### 4.4 实验结果与分析

各模型在测试集上的性能对比如下表所示：

表 1 不同变压器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)	0.901	0.72	0.65	0.68
随机森林(RF)	0.938	0.81	0.78	0.79
梯度提升树(GBDT)	0.947	0.83	0.82	0.82
长短时记忆网络(LSTM)	0.945	0.85	0.84	0.85

结果分析：

整体上，所有模型都取得了不错的预测效果，证明了基于 DGA 时序数据进行故障预测的可行性。

树模型表现优于线性模型，说明变压器故障与气体特征之间存在复杂的非线性关系。

LSTM 模型在精确率、召回率和 F1-Score 上均达到最优，特别是 F1-Score 达到了 0.85。这表明 LSTM 能够有效学习 DGA 数据中的时序动态特性，更准确地识别出真正的故障前兆，同时保持较低的误报率。这对于实际运维中减少不必要的恐慌性停电检修具有重要意义。

对 RF 和 GBDT 模型进行特征重要性分析，发现“C<sub>2</sub>H<sub>2</sub>近 30 天斜率”、“C<sub>2</sub>H<sub>2</sub>/C<sub>2</sub>H<sub>4</sub>比值”和“H<sub>2</sub> 绝对含量”是排名最前的特征，这与电工理论中乙炔和氢气是反映放电和过热的关键气体的认知高度一致，增强了模型的可解释性。

为了更直观地展示 LSTM 模型的预测能力，选取了一台最终发生故障的变压器，绘制了其故障前一年

内，模型预测的故障风险概率变化曲线，如图 2 所示。可以清晰地看到，在故障发生前约 4 个月，风险概率开始波动上升，并在故障前 1 个月持续超过 0.7 的高风险阈值，成功实现了早期预警。

#### 5 挑战与展望

尽管大数据分析技术在变电运维中展现出巨大潜力，但其广泛应用仍面临挑战：

**数据质量难题：**监测数据存在噪声、缺失、不一致等问题，对数据治理提出了高要求。

**模型可解释性：**深度学习等复杂模型如同“黑箱”，其决策过程难以理解，影响运维人员的信任和采纳。发展可解释性 AI 是重要方向。

**跨域数据融合：**如何有效融合电气、机械、化学、环境等多物理场数据，以及文本、图像等多模态数据，是实现更精准诊断的关键。

**边缘与云端协同：**需研究分析任务在边缘侧和云端的合理分配与协同机制。

未来，随着数字孪生技术的成熟，构建与物理变电站实时同步的虚拟模型，并在此基础之上进行仿真、预测和优化，将成为智能变电运维的重要发展趋势。同时，结合知识图谱技术，将专家经验与数据驱动模型深度融合，也将进一步提升运维决策的智能化水平。

#### 6 结论

本文系统研究了大数据分析技术在智能电网变电运维中的应用。通过构建多层次的技术框架，深入分析了其在状态评估、故障预测、运维优化等核心场景下的实现路径。针对变压器故障预测的具体实验表明，基于 LSTM 等深度学习算法能够有效挖掘 DGA 时序数据中的深层规律，实现高精度的早期故障预警，其性能优于传统机器学习方法。研究成果证实了大数据分析技术是推动变电运维向智能化、精准化、高效化转型升级的核心驱动力，具有重大的理论价值和广阔的应用前景。

#### 参考文献：

- [1] 肖湘宁, 罗金山, 陶顺.智能电网大数据技术研究现状与发展展望[J].电力系统自动化, 2019,43(1):2-15.
- [2] 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.
- [3] Zhao W, Zhang C, Li Y, et al. A novel transformer fault diagnosis model based on optimal wavelet neural network and dissolved gas analysis[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2020,27(4):1349-1357.
- [4] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997,9(8):1735-1780.
- [5] 孙宏斌, 郭庆来, 王彬.能源互联网大数据分析研究综述与展望[J].中国电机工程学报, 2022,42(1):1-20.