International Journal of Grid and High Performance Computing
Volume 17 • Issue 1 • January-December 2025

Transformer Fault Diagnosis Based on Parallel AdaBoost-NB Algorithm on Spark Cloud Platform

Cheng Liu
https://orcid.org/0009-0004-4541-2161
Jiangsu Vocational Institute of Commerce, China

Lin Ji
https://orcid.org/0009-0008-8325-7755
Jiangsu Vocational Institute of Commerce, China

ABSTRACT

Power transformers are crucial equipment in the power grid because they are essential for ensuring stable grid operation. Sequential machine learning and artificial intelligence diagnostic algorithms often face issues of low efficiency and prolonged processing times with respect to handling large volumes of oil-immersed transformer fault data. In this article, the authors propose a new transformer fault diagnosis method that is based on the parallel AdaBoost-Naive Bayes algorithm. This method allows for resampling and reweighting, making the model pay more attention to samples that are difficult to classify and thereby improving performance on imbalanced datasets. The Spark platform is used for parallel processing of massive data, utilizing the cluster's multiple nodes for efficient fault diagnosis. Experimental results show that compared with traditional diagnostic methods, the proposed method achieves a significant improvement in diagnostic accuracy, with an accuracy rate of 93.38%. The significant speedup ratio achieved by parallel processing technology underscores its effectiveness and advantages in handling large-scale transformer fault data.

KEYWORDS

Power Transformer, Fault Diagnosis, AdaBoost Algorithm, Naive Bayes Algorithm, Spark

INTRODUCTION

Transformers are crucial pieces of equipment in the power system, and their normal operation is vital to ensuring the system's stability and power supply. However, because of long-term operation under high voltage and high current conditions, they have a high failure rate, and their malfunction can severely affect the power system's stability. Therefore, the fault diagnosis of power transformers plays a significant role in the reliability and safety of modern power systems, making this a hot topic in both industrial and academic research. Given their unpredictability, it is necessary to quickly locate and isolate faults to minimize their impact on transformers. Effective fault diagnosis of power transformers can prevent expensive repairs, downtime, and personnel hazards, and it helps avoid damage to nearby equipment (Abbasi, 2022). At present, transformer fault diagnosis methods mainly include visual inspection, electrical preventive testing, dissolved gas analysis (DGA) in oil, expert systems (Wu et al., 2024), artificial neural networks (Zhang et al., 2024), intelligent systems, diagnosis technology based

DOI: 10.4018/IJGHPC.381295

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

on acoustic signature, and fault diagnosis based on deep learning, among others. Visual inspection preliminarily judges faults by observing the external characteristics of the transformer, including abnormal conditions like temperature, sound, and vibration. Electrical preventive testing assesses the condition and performance of the transformer through methods such as insulation resistance measurement. Diagnosis technology based on acoustic signature extracts sound characteristics during the operation of the transformer and uses pattern recognition technology for fault diagnosis. Expert systems and artificial neural networks use expert knowledge and experience, combined with artificial intelligence technology, to diagnose and analyze faults in transformers. Intelligent systems integrate various intelligent technologies, such as machine learning and deep learning, to analyze the operation data of transformers and achieve intelligent fault diagnosis.

In identifying faults in power transformers, the method based on DGA offers advantages, such as being unaffected by external electromagnetic environments, has easy online implementation, and has the capability to accurately determine potential fault types. It also predicts potential faults and defects in transformers, thus enabling preemptive measures. Therefore, it is an effective method for fault identification in power transformers (Zhou et al., 2019). Although DGA in oil diagnoses internal faults in transformers, such as partial discharges or overheating, by analyzing the composition and content of gases dissolved in the oil, it is a critical technique for power system maintenance. Traditional DGA diagnostic methods, such as the three ratios and Rogers' methods, are rule based and straightforward to implement; however, their applicability to new or complex fault patterns is limited by their reliance on predefined rules (Tokunaga et al., 2017). To further improve fault diagnosis accuracy, researchers have adopted machine learning algorithms, such as the support vector machines (SVM) algorithm. The SVM algorithm performs well and has strong learning generalization capabilities and high classification accuracy when dealing with limited or small datasets. However, its performance significantly declines when DGA samples increase (Li et al., 2016). Some researchers have focused on optimizing the kernel function parameters in the SVM algorithm, using genetic algorithms and particle swarm optimization to enhance the model's generalization capability (Zheng et al., 2011, 2014; Xue et al., 2015). In addition, studies of the application of android malware detection (AlSobeh et al., 2024a) have adopted a time-aware machine learning framework. This time-feature–based analysis method provides a new perspective for transformer fault diagnosis. Some researchers (AlSobeh et al., 2024b) have used large language models and statistical learning to enhance runtime monitoring, and the advantages of artificial intelligence and deep learning technologies are quite obvious. Other scholars have tackled issues such as insufficient and unbalanced fault samples and limited gas data features using deep-coupled dense convolutional neural networks (Li et al., 2022).

To address the inefficiency of convolutional neural networks and recurrent neural networks in extracting long-term dependency features, some researchers have proposed a modified transformer model, the target transformer, which features self-attention and target-attention mechanisms for chemical process fault diagnosis (Wei et al., 2022). To address time-delay issues in DGA sequence predictions of deep learning methods, some researchers have introduced a novel hybrid model that combines hierarchical attention networks and recurrent long short-term memory networks to eliminate this phenomenon (Zhong et al., 2023).

Several recent studies have further expanded the research on transformer fault diagnosis. Wu et al. (2019) proposed a transformer fault identification method that is based on a self-adaptive extreme learning machine, which leverages the machine's adaptability to handle complex data and improve diagnostic accuracy. Liu et al. (2024) integrated the improved quantum-behaved particle swarm optimization with random forests for fault diagnosis, optimizing model parameters to enhance its overall performance and robustness. Yao et al. (2024) introduced a digital twin model to the field of oil-immersed transformer fault diagnosis, enabling real-time simulation of transformer operation status and more precise fault prediction. Kumar et al. (2024) focused on optimizing the fault classification of dissolved gases in transformers by using data synthesis and dimension-reduction techniques, effectively addressing data insufficiency and improving classification efficiency. These studies not only represent

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/transformer-fault-diagnosis-based-on-parallel-adaboost-nb-algorithm-on-spark-cloud-platform/381295

Related Content

Service and Management Oriented Traffic Information Grid
Yu Fang, Dong Liang Zhang, Chun Gang Yan, Hong Zhong Chenand Changjun Jiang (2012). *Technology Integration Advancements in Distributed Systems and Computing (pp. 283-295).*
www.irma-international.org/chapter/service-management-oriented-traffic-information/64454

Credential Management Enforcement and Secure Data Storage in gLite
Francesco Tusa, Massimo Villariand Antonio Puliafito (2012). *Grid and Cloud Computing: Concepts, Methodologies, Tools and Applications (pp. 956-978).*
www.irma-international.org/chapter/credential-management-enforcement-secure-data/64524

A New Social Volunteer Computing Environment With Task-Adapted Scheduling Policy (TASP)
Nabil Kadacheand Rachid Seghir (2021). *International Journal of Grid and High Performance Computing (pp. 39-55).*
www.irma-international.org/article/a-new-social-volunteer-computing-environment-with-task-adapted-scheduling-policy-tasp/273653

EcoGrid: A Toolkit for Modelling and Simulation of Grid Computing Environment for Evaluation of Resource Management Algorithms
Hemant Kumar Mehta (2014). *International Journal of Grid and High Performance Computing (pp. 1-16).*
www.irma-international.org/article/ecogrid/119449

Measuring the Characteristics of Hypervisor I/O Scheduling in the Cloud for Virtual Machine Performance Interference
Ziye Yang, Haifeng Fang, Yingjun Wuand Chunqi Li (2013). *International Journal of Grid and High Performance Computing (pp. 5-29).*
www.irma-international.org/article/measuring-the-characteristics-of-hypervisor-io-scheduling-in-the-cloud-for-virtual-machine-performance-interference/102754