

基于 XGBoost 的电网过流异常检测模型^{*}

马智强¹ 康洁滢¹ 梁飞¹ 王进²

(1. 国网宁夏电力有限公司营销服务中心(国网宁夏电力有限公司计量中心) 宁夏 750010; 2. 成都大学 成都 610106)

摘要: 在电网系统中,异常检测效率直接影响到系统维护成本,传统电网异常检测方法基于专家经验知识转化为固定规则与阈值,存在一定的局限性。现有的异常检测研究多以窃电、设备故障为主要分析对象,对过流异常的分析不足。本文针对过流异常的特性,分析了传统经验规则存在的问题与缺陷,通过特征工程确定了特征量,提出了基于 XGBoost 的电网过流异常检测模型。通过实验数据测试与评估,本文模型在 5 折交叉验证中 F1 分数最低值相较于传统规则提升了 19.2%,平均值相较于传统规则提升了 15.1%,各项实验指标均优于基于传统经验规则的检测方法,且没有出现明显的性能差异,证明了模型的检测效果。与异常检测常用的其他机器方法对比,本文模型的 F1 分数提升了 6.4% 至 8.7%,稳定性及准确性均有优势。通过训练数据远少于测试数据的极端情况测试以及对模型进行的可解释性分析表明,本文模型具有较高的透明度、可信度,同时具有良好的泛化性能,可以有效支撑在实际环境中推广应用过流异常检测。

关键词: 过电流; 异常检测; 机器学习; 可解释性

中图分类号: TM73; TN06 **文献标识码:** A **国家标准学科分类代码:** 470.4054

XGBoost-based power grid overcurrent anomaly detection model

Ma Zhiqiang¹ Kang Jieying¹ Liang Fei¹ Wang Jin²

(1. State Grid Ningxia Electric Power Co., Ltd. Marketing Service Center (State Grid Ningxia Electric Power Co., Ltd., Metering Center), Ningxia 750010, China; 2. Chengdu University, Chengdu 610106, China)

Abstract: Traditional power grid anomaly detection methods rely on converting expert knowledge into fixed rules and thresholds, which cannot meet the demands of rapidly evolving power grid systems. The current anomaly detection research mainly focuses on electricity theft and equipment failures as the main analysis objects, but the analysis of overcurrent anomalies is insufficient. This paper analyzes the characteristics of overcurrent anomalies, and discusses the problems and deficiencies of traditional experience-based rules. Through feature engineering, we determines the feature variables, and proposes an XGBoost-based power grid overcurrent anomaly detection model. Through experimental data testing and evaluation, the indicators of the model proposed in this paper outperform the detection methods based on traditional experience-based rules. In the 5-fold cross-validation, the minimum F1 score of the proposed model showed a 19.2% improvement compared to traditional rules, while the average value demonstrated a 15.1% improvement. The experimental results did not show significant performance differences, confirming the effectiveness of the model in anomaly detection. Compared to other commonly used machine methods for anomaly detection, the proposed model in this paper achieved an improvement of 6.4% to 8.7% in F1 score, demonstrating advantages in terms of stability and accuracy. The extreme case testing with training data significantly less than the testing data, along with the conducted interpretability analysis of the model, demonstrated that the proposed model exhibits high transparency and reliability. Moreover, it shows good generalization performance, making it suitable for effective deployment in real-world environments for overcurrent anomaly detection.

Keywords: overcurrent; anomaly detection; machine learning; interpretability

0 引言

21 世纪以来,随着电网系统的飞速发展,异常检测变

得越来越重要。及时、准确地检测出电网异常状态,是维持电网系统稳定、健康运行不可或缺的重要环节。随着智能化电网建设发展以及数据存储技术发展,目前电网系统采

集并保存了海量的监测数据^[1],使电力数据分析不再局限于现场采集测试,机器学习在分析电网数据中开始广泛应用^[2]。

目前国内外学者对电力数据异常检测方法研究主要分为两大类^[3]:1)基于专家经验知识的方法,将专家的异常检测知识与长期积累的经验转化为固定的规则与阈值,具有很强的主观性;2)基于机器学习的方法^[4],对电网检测数据分析预测,从而完成异常检测。主要分为 K 近邻算法(k-nearest neighbors, KNN)^[5]、支持向量机(support vector machine, SVM)^[6]、随机森林^[7]等,这类方法的特征选择直接影响异常检测的最终效果,需要针对具体数据人工设计特征模型。文献[5]提出了一种加权 KNN 数据异常检测方法,适用智能电网中的高维数据;文献[6]针对坏数据、负荷突变和单相接地等类型的异常,提出了一种基于支持向量机的多类型异常检测方法。上述机器学习方法多针对电力系统自身异常故障,对用户超容过流用电导致的用户侧异常研究较少。

在各种异常类型中,过流异常不同于其他异常故障,电网系统本身具有一定的过电流能力^[8],短时间、非极大值的过流对电网影响很小,无需特别处理,而持续过流或者突发极大值过流会导致过度发热,存在火灾或设备损坏风险。基于这种特性,为了充分利用设备自身特性,减少额外的检修人力资源浪费,电网系统中仍使用基于专家经验知识的方法,设置特定规则来检测异常。由于电网数据爆发式增长^[9],规模不断扩大,传统基于经验知识的规则与固定阈值主观性较强,通用性较为薄弱,存在一定的虚报漏报现象。针对以上问题,文中以实例分析了传统过流异常检测规则局限性,引入极致梯度提升树(extreme gradient boosting, XGBoost)算法^[10],建立电网过流异常检测模型,优化机器学习预测效果。最后,基于真实案例数据,通过实验对比,验证了文中模型在电网过流异常检测时的可靠性和有效性。

1 电网过流异常状态检测分析

目前过流异常检测规则为出现连续 1 h(每日 96 点曲线中连续 4 个点)电流数值大于电能表最大电流或互感器额定电流时,记为过流异常。该规则基于传统经验知识,在多数情况下可以较好地判断过流异常,但随着电网规模越来越大,不同用户的用电情况各不相同,仍有部分异常并不能准确判断。

根据某市用户过流异常统计及每日电流曲线数据,可以发现存在部分用户存在电流数据虚报漏报现象,如图 1、2 所示。在图 1(a)中,该用户接入类型为直接接入式,过流阈值为 60 A,可以看到该户在某一天实际存在持续过流情况,但因传统经验规则连续 4 点判断限制,并未报告异常,存在安全隐患;图 1(b)中,该用户某一天出现了若干次极大值,远远超过设定阈值 60 A,同样并未记为异常。

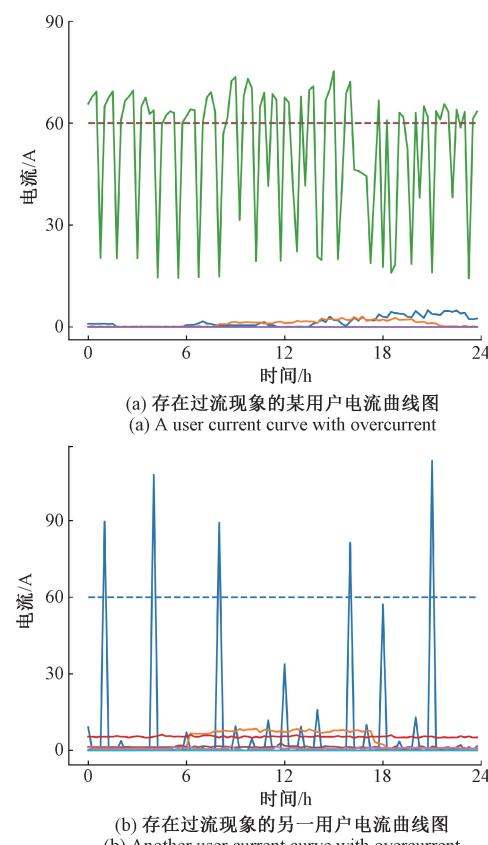


图 1 传统规则下漏报电流曲线情况
Fig. 1 A user's omission reporting under traditional rules

如图 2 所示,该用户某日内短时间达到过流阈值,但其前后长时间内并未再出现过流现象,这类情况可以不用额外处理,利用电力设备自身的过电流能力^[11],节约检修人力资源。

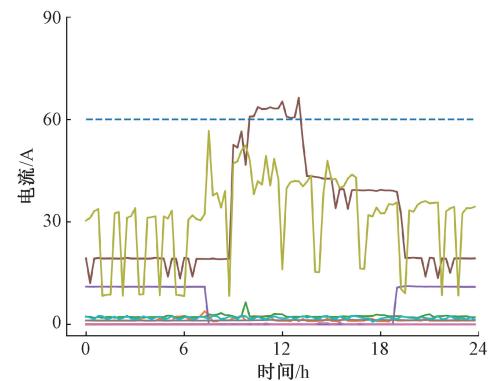


图 2 传统规则下虚报电流曲线情况
Fig. 2 A user's false reporting under traditional rules

通过现有数据分析发现,在大部分的过流异常情况下,用户都存在三相电流不平衡问题,往往出现过流情况的某相电与其他两相存在较大差值,如图 3 某用户三相电过流情况所示。为了提升模型泛化性能,挖掘数据潜在规律,在

检测过流异常时,可以考虑引入三相不平衡度参数作为特征量。

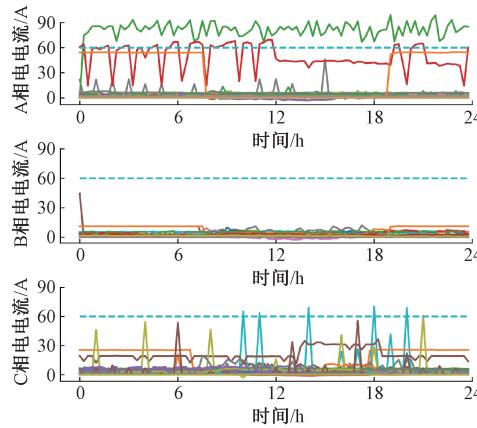


图 3 某用户三相电过流情况

Fig. 3 A user's three-phase power overcurrent situation

2 电网过流异常检测模型

本文构建电网过流异常检测模型流程分为特征工程、模型分析与训练、模型评估、可解释性分析 4 个部分。

2.1 特征工程

模型特征指机器学习模型在训练和推断过程中用来进行预测的输入。特征工程是机器学习流程中重要的组成部分,是从原始数据中提取和转换变量作为模型特征的过程。良好的特征工程可以降低模型复杂度,提高模型性能。本文特征工程主要分为数据预处理、特征选择两部分:

1)数据预处理:现有电网数据为数据库中提取,分为普通用户每日电流曲线数据和已记录的过流异常用户详细信息。数据中存在部分缺失项,以及例如户名等文字类型数据、户号等无具体数值大小意义的数字,需要通过处理后才能开始训练。对电网数据的预处理主要包括缺失数据处理和数据转换:对电流数据中的缺失值,参考现有用电采集系统研判规则,采用以下方法补全:当连续时间内缺点数小于等于 2 个点时,取该时间点前 4 日的平均值补全;当连续时间内缺点数大于 2 个点时,取该点前 2 日的电流数据曲线近似拟合补全。对于单日缺失数据超过 1/4(24 个点)的数据,作删除处理,对三相电表的用户数据中 A、B、C 相电流单独作为一条数据。对用户类型、接入类型等离散型特征,采用独热编码^[12]处理,可以避免标签编码中的数值假设问题^[13]影响训练效果。

2)特征选择:由于电网数据较为详细,部分指标与异常判断无明显关联,无关的特征指标输入会消耗训练资源,增大模型误差。因此,首先需要利用相关性分析^[14]筛选出与电流信息关联性强的指标,合理选择特征,剔除无关数据。通过上节数据分析,在模型训练时将加入三相不平衡度^[15]特征,实验时与基准性能进行对比,观察该特征量对基准性能的影响。

2.2 基于 XGBoost 的电网过流异常检测模型

XGBoost 是梯度提升决策树^[16] (gradient boosting decision tree, GBDT) 算法的一种高效实现。XGBoost 使用决策树作为基础模型,核心是 Boosting 思想,多个决策树共同决策,后一棵决策树的生成将之前所有树的偏差考虑在内,以此达到整个模型效果提升。XGBoost 模型可以定义为:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (1)$$

式中: F 代表分类回归树(classification and regression tree, CART) 空间, f_k 为基学习器, $f_k(x_i)$ 代表第 k 棵树对样本 x_i 的预测分数, \hat{y}_i 表示预测结果。

XGBoost 模型的目标函数定义如下:

$$Obj(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (3)$$

式中: $l(y_i, \hat{y}_i)$ 表示第 i 个样本预测偏差; $\Omega(f)$ 表示模型复杂度; T 表示决策树数量, ω 表示权重, γ 和 λ 为超参数。

设 $\hat{y}_i^{(t)}$ 为第 t 次迭代时预测第 i 个样本结果,则 t 棵树的目标函数为:

$$Obj^t(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i)) + \Omega(f_t) \quad (4)$$

对公式二阶泰勒展开可以得到:

$$Obj^t(\theta) \approx \sum_{i=1}^n [l(y_i, \hat{y}_i^{(t-1)}) + (g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i))] + \Omega(f_t) \quad (5)$$

式中: g_i 和 h_i 为损失函数的一阶偏导数和二阶偏导数,且 $l(y_i, \hat{y}_i^{(t-1)})$ 为常数项,可将目标函数表示为:

$$Obj^t(\theta) = \sum_{j=1}^T [(\sum_{i \in I_j} g_i) \omega_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) \omega_j^2] + \gamma T \quad (6)$$

对式(5)求导,导数为 0 时可得到叶子结点权重 ω_j^* :

$$\omega_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (7)$$

将 ω_j^* 带入原目标函数即可得到新的目标函数:

$$Obj^t = -\frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (8)$$

2.3 模型评估

文中对模型效果评估采用 5 折交叉验证的方式,将数据集随机分为 5 份数量相同的子集,轮流将其中一份作为测试集,剩余 4 份作为训练集。交叉验证可以减少因样本划分不同而引入的差别,使模型评估更加有效。

由于在实际应用中会出现训练数据会远少于测试数据的情况^[17],为了更准确评估模型效果,实验时还将通过更改数据集中训练集比例,模拟实际应用出现的极端情况,训练集取值范围为[0.1,0.9],取值间隔0.1。

2.4 可解释性分析

可解释性指的将模型的预测过程转化成具备逻辑关系的规则的能力^[18],即人类能够理解模型预测原因的程度。在现实场景中,机器学习模型检测效果很好并不意味着可以直接投入业务应用,在不了解模型是如何使用特征进行预测的情况下,是无法对模型产生信任并对齐有效管理优化的。对模型进行可解释性分析,可以提高模型的可信度和透明度,对模型的泛化性能提供有力支持。

本文采用 SHAP(Shapley additive explanations)对模型进行可解释性分析。SHAP 是一种模型无关的机器学习解释方法^[19],基于博奕论中 Shapley 值的概念,为每个输入特征生成一个值,该值可以表示出这个特征如何参与预测过程,如对预测结果产生积极或是负面影响等。

Shapley 值是合作博奕论中的一个概念,它提供了一种在参与贡献的成员之间公平分配支出的数学方法。Shapley 值的形式化定义为:设 N 为 n 个成员的集合, S 为 N 的任意子集, $v(S)$ 为各子集的获利,当任何情况下合作比单人贡献要更有利时,每个人获得的利益 $\varphi_i(v)$ 可以表示为:

$$\varphi_i(v) = \sum_{S \in N} \frac{[(|S|-1)! (n-|S|)!]}{n!} [v(S) - v(S \setminus \{i\})] \quad (9)$$

式中: $v(S) - v(S \setminus \{i\})$ 称为成员 i 在参与合作 S 中的边际贡献,乘上加权因子后,即为成员 i 的 Shapley 值。

3 实验分析

3.1 数据集介绍

本文中所有数据均来自某地区电网系统智能电表自动采集后提取。由于目前过流异常检测流程为电网系统中记录到异常数据后,派发工单由专业人员去现场确认是否真实存在异常现象,故本文数据集采集用户范围为 2024 年 1 月~2024 年 3 月该地区派发的异常工单明细中涉及到的所有用户,以及同区域、同用电类型的随机抽取部分用户,记录时间为 2024 年 1 月~2024 年 3 月。过流异常明细记录表包含 103 条记录,57 户不重复用户,涵盖了高压与低压用户、直接接入式与经互感器接入式两种接入电网方式;每日电流曲线表中包含 28 700 条记录,113 户不重复用户(包括过流异常明细表中记录的 57 户)。将所有现场核查为过流异常的用户异常时间段数据标记为异常,其余数据标记为正常。经过数据预处理后,得到过流数据 1 181 条,正常数据 27 090 条,表 1 为数据预处理后的数据集概况,其中记录日期数含义为:用户单日电流数据记一次,不区分三相电流数据:

表 1 数据集概况
Table 1 Overview of dataset

类型	样本总数	用户数	记录日期数
过流	1 181	57	103
正常	27 090	56	10 067
总计	28 271	113	10 170

在原始数据中,如管理单位编号与管理单位名等数据互为冗余参数,实际意义没有区别,因此去掉文字类的地址与名称信息;编号本身无数字意义,为防止编号数值太大影响模型训练效果,将各类编号与其他离散属性均通过独热编码后作为特征量。将计算得到的三相电流不平衡度参数作为特征量。实验表明,加入三相电流不平衡度特征后,模型效果得到了提升。图 4 为原始数据与加入三相电流不平衡特征后模型效果对比:

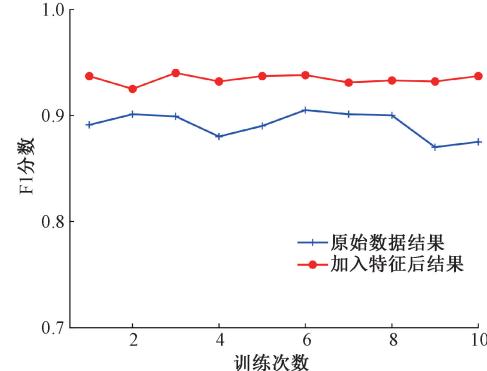


图 4 加入特征前后训练效果对比
Fig. 4 Comparison of training effects before and after adding features

图 5 为预训练模型特征重要性图(部分),为了直观比较其他特征,图中略去每日 96 点电流数据特征。由图 5 可以看出,管理单位编号、数据类型、单位编号、设备编号及用户编号特征重要性值较低。考虑到数据样本实际情况以及模型训练效果,最终选取三相电流不平衡度、接入类型和每日电流数据作为特征输入。下文其他算法模型对比以此特征输入为参考。

3.2 测试基准

测试时根据传统经验规则分析数据,统计出传统规则下检测效果,然后分别与异常检测中几种常用机器学习分类算法进行比较,包括 KNN、SVM、随机森林^[20]。确定相关参数范围^[21]后,使用网格搜索法获得最优参数,具体参数如下:

1) KNN^[22]: KNN 的核心思想是如果一个样本在特征空间中的 K 个最近邻的样本中的大多数属于某一类别,则该样本也属于这个类别。参考文献[22],超参数 K 值($n_{neighbors}$)取值范围为[1,10],取值间隔为 1; 距离度量 P (metric)的选择范围为: 欧氏距离、曼哈顿距离、切比雪夫

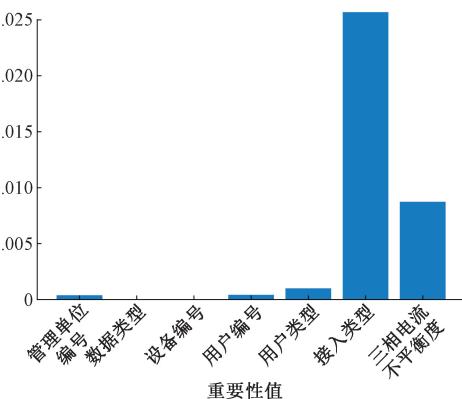


图 5 部分特征重要性值

Fig. 5 Partial feature importance value

距离、闵可夫斯基距离。最终确定 K 值为 5, 距离度量为闵可夫斯基距离。

2) SVM^[23]: SVM 的基本原理是寻找一个分割, 将样本空间一分为二。参考文献[23], 惩罚参数取值范围为 [0.001, 0.01, 0.1, 1, 10, 100]; 核函数取值范围: 线性核函数、多项式核函数、高斯核函数。最终确定惩罚参数 10, 径向基核函数。

3) 随机森林^[24]: 随机森林是一个包含多个决策树的分类器, 其输出的类别是由个别树输出类别的众数决定。参考文献[24], 超参数决策树的取值范围为 [100, 500], 取值间隔为 100; 树的最大深度为 [10, 30], 取值间隔 10。最终确定树的数量为 300, 最大深度为 10。

3.3 评价指标

文中以 F1 分数^[25]来评价过流异常分析方法。F1 分数是统计学中用来衡量二分类模型精确度的一种指标。F1 分数是精准率和召回率的调和平均数, 指标定义如下:

$$P = \frac{TP}{TP + FP} \quad (10)$$

$$R = \frac{TP}{TP + FN} \quad (11)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (12)$$

式中: P 表示精准率(查准率), R 表示召回率(查全率), TP 为真阳性数量(实际意义为正确检测到的过流样本数); FP 假阳性数量(实际意义为未被检测出的过流样本数); FN 表示假阴性数量(实际意义为被检测为异常的正常样本数)。

3.4 过流异常检测模型结果分析

图 6 为实例数据中传统规则经过统计计算后与文中模型实验结果对比。

由图 6(a)可以看出, 在 5 折交叉验证情况下, 传统规则 F1 分数最低为 0.767, 最高为 0.857, 平均值为 0.81; 而 XGBoost 模型 F1 分数最低为 0.914, 最高为 0.95, 平均值为 0.932, 最低值相较于传统规则提升了 19.2%, 平均值相较于传统规则提升了 15.1%。同时, 在 5 折交叉验证中各

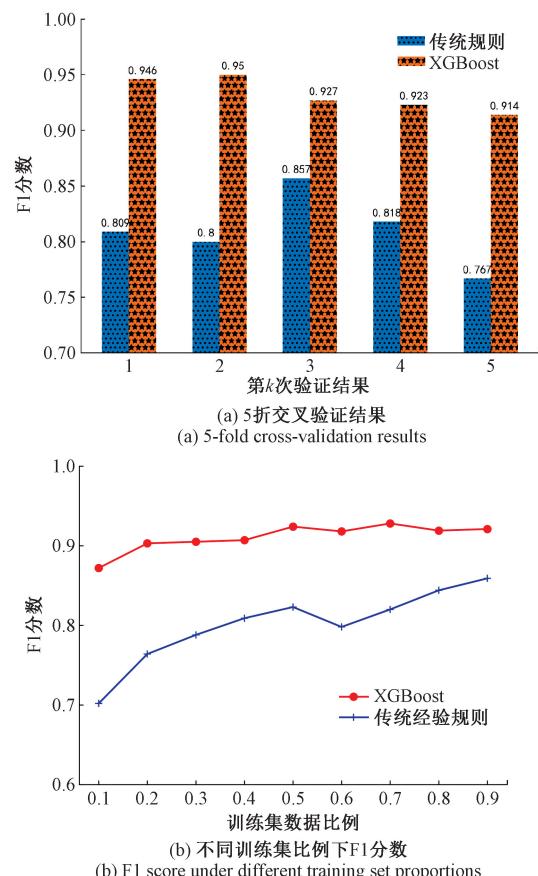


图 6 传统经验规则与文中模型实验结果对比

Fig. 6 Comparison of traditional empirical rules and model experimental results

分数没有明显差异, 说明该模型在不同的测试集与训练集情况下具有良好的泛化性能。

由图 6(b)可以看出, 在训练集比例为 0.1 时, 传统规则 F1 分数明显变差, 仅为 0.702; 而 XGBoost 模型在不同训练集数据比例下整体较为稳定, 训练集比例为 0.1 时 F1 分数仍有 0.872, 相较于传统规则提升了 24.2%, 说明在训练集数据极少情况下, XGBoost 模型对过流异常的检测效果明显好于传统经验规则。

图 7 为通过上文网格搜索法确定最优超参数后, 其他常用机器学习算法模型对比文中模型效果。经过计算, KNN 模型在 5 折交叉验证中平均 F1 分数为 0.876, SVM 模型平均 F1 分数为 0.857, 随机森林模型平均 F1 分数为 0.863。测试效果表明, 在 5 折交叉验证结果以及不同训练集比例下实验, XGBoost 模型效果相较其他常用机器学习算法, 提升了 6.4%~8.7%。

图 8 为通过 SHAP 对模型可解释性分析的特征贡献度绝对值蜂窝图, 因 96 点电流数据贡献程度接近, 为了便于直观展示, 图中仅展示 2 个点电流数据用以对比, 其余电流数据折叠显示。由图可以看出, 在所有输入特征中, 电流数据对预测贡献程度最高, 说明电流数值大小为判断是否

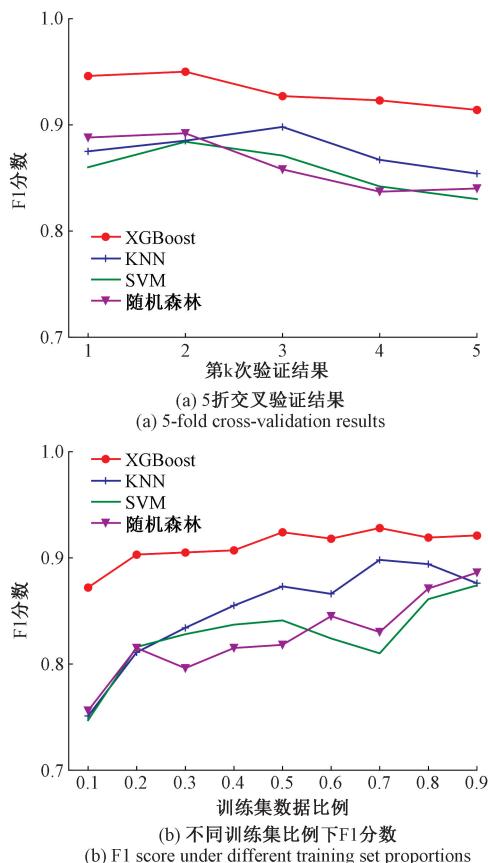


图 7 不同机器学习算法实验结果对比

Fig. 7 Comparison of experimental results from different machine learning algorithms

过流的最基本依据;接入类型特征对预测贡献程度仅次于电流数据总和,大于单点电流数据,而现实中接入类型不同,变压器的过电流能力也不相同,在传统经验规则中,接入类型直接决定了判断过流异常的阈值,因此模型判断时该项特征重要程度较高。

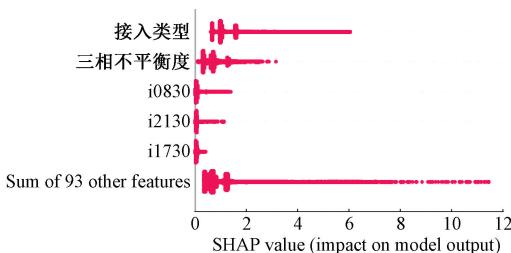


图 8 SHAP 特征贡献程度分析蜂窝图

Fig. 8 SHAP feature contribution analysis in the form of a heatmap

图 9 为单个样本 SHAP 分析图,图中特征名左侧数值为该特征实际值,“i0130”表示 01:30 时刻的电流采集值,下同。该样本被模型检测为过流异常。该图直观表示了模型做出预测的原因,每个特征量对预测结果的影响程度,SHAP 值为正表示该特征量对预测结果为正面影响。其中:接入类型对判断结果影响最大,接入类型 1 表示该样本

为经互感器接入式用户;三相不平衡度达到了 0.803;多个点电流数据超过了 6 A。结合所有特征输入后,模型检测结果为异常。

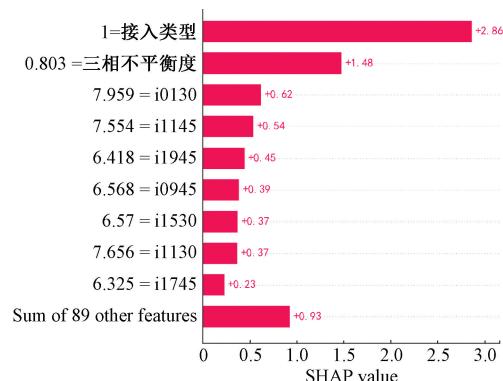


图 9 检测为过流异常的某样本 SHAP 分析

Fig. 9 SHAP analysis of a sample identified as an overcurrent anomaly

由可解释性分析可以看出,该模型做出决策的主要参考特征基本符合经验知识,表示模型可信程度较高;各个特征量没有出现明显异常的重要性值,表示模型具有一定的泛化性能,可以应用于实际场景。

4 结 论

针对电网数据中过流异常检测的需求,本文分析了过流异常的特性,实例验证了传统过流异常检测经验规则的不足,通过特征工程确定了特征量,引入了基于 XGBoost 的电网过流异常检测模型。通过实验结果分析,在真实数据集中,文中模型测试效果优于传统经验规则检测方法及其他常用机器学习分类算法。交叉验证结果显示文中模型过流异常检测性能稳定、准确,在不同数据集与测试集中体现了良好的泛化性能;不同测试集数据比例实验结果显示,文中模型在测试数据与训练数据极端不平衡情况下,仍具有良好的检测效果。可解释性分析结果表明,文中模型透明度与可信度较高,没有明显的异常特征比重,表示具有一定的泛化性能,可以有效支撑在实际环境中代替传统规则检测过流异常。

参考文献

- [1] 张译, 张广德, 于连城, 等. 基于数据仓库的电网数据分析系统的设计与开发 [J]. 电子测量技术, 2018, 41(15): 35-39.
ZHANG Y, ZHANG G D, YU L CH, et al. Design and development of grid data analysis system based on data warehouse [J]. Electronic Measurement Technology, 2018, 41(15): 35-39.
- [2] 周念成, 廖建权, 王强钢, 等. 深度学习在智能电网中的应用现状分析与展望 [J]. 电力系统自动化, 2019, 43(4): 180-191.

- ZHOU N CH, LIAO J Q, WANG Q G, et al. Analysis and prospect of deep learning application in smart grid [J]. Automation of Electric Power Systems, 2019, 43(4): 180-191.
- [3] 白开峰, 赵宏斌, 张芸, 等. 电网异常业务数据检测方法综述[J]. 计算机与现代化, 2023(3): 79-83, 89.
- BAI K F, ZHAO H B, ZHANG Y, et al. Review of abnormal service data detection methods in power grid[J]. Computer and Modernization, 2023(3): 79-83, 89.
- [4] 何成艳, 刘姜, 刘丽婕. 电网环境下基于深度学习的推荐系统算法研究[J]. 电子测量技术, 2020, 43(12): 60-64.
- HE CH Y, LIU J, LIU L J. Research on recommendation algorithm based on deep learning in power grid environment[J]. Electronic Measurement Technology, 2020, 43(12): 60-64.
- [5] 林舒嫄, 林晓敏, 欧亚, 等. 基于 kNN 算法的智能电网 5G 海量接入数据异常检测[J]. 粘接, 2024, 51(2): 155-158.
- LIN SH Y, LIN X M, OU Y, et al. Abnormality detection of 5G massive access data of smart grid based on kNN algorithm [J]. Adhesion, 2024, 51(2): 155-158.
- [6] 郭嘉辉, 侯月婷, 丁磊, 等. 基于支持向量机的电力系统状态估计多类型数据异常检测[J]. 国外电子测量技术, 2024, 43(4): 152-161.
- GUO J H, HOU Y T, DING L, et al. Multi-type data anomaly detection in power system state estimation using support vector machine[J]. Foreign Electronic Measurement Technology, 2024, 43(4): 152-161.
- [7] 王金忠, 吴焰龙. 基于随机森林的智能电网多源数据异常检测[J]. 电子设计工程, 2024, 32(7): 149-152, 157.
- WANG J ZH, WU Y L. Multi-source data anomaly detection in smart grid based on random forest[J]. Electronic Design Engineering, 2024, 32(7): 149-152, 157.
- [8] GUPTA K, SARANGI S. Adaptive overcurrent relay setting for distribution system using superconducting fault current limiters[C]. 2018 IEEE 8th Power India International Conference(PIICON), 2018: 1-6.
- [9] ZHENG Z B, YANG Y T, NIU X D, et al. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids[J]. IEEE Transactions on Industrial Informatics, 2018, 14(4): 1606-1615.
- [10] LIU Y H, LUO H, ZHAO B, et al. Short-term power load forecasting based on clustering and XGBoost method [C]. 2018 IEEE 9th International Conference on Software Engineering and Service Science(ICSESS), 2018: 536-539.
- [11] 黄景光, 丁婧, 郑淑文, 等. 基于电流突变量的自适应过电流保护新原理[J]. 电力系统保护与控制, 2018, 46(7): 49-55.
- HUANG J G, DING J, ZHENG SH W, et al. A new adaptive over current protection principle based on current mutation[J]. Power System Protection and Control, 2018, 46(7): 49-55.
- [12] WANG H D, GUPTA G. Superfast selection for decision tree algorithms[J]. ArXiv preprint arXiv: 2405.20622, 2024.
- [13] HANG J Y, ZHANG M L. Collaborative learning of label semantics and deep label-specific features for multi-label classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(12): 9860-9871.
- [14] TRITHAM C, LEKAWAT L, ARRAYANGKOOL A, et al. A comparison between correlation and grey relational for big data and analytics [C]. 2018 International Electrical Engineering Congress, 2018: 1-4.
- [15] 吴俊, 诸军, 沈海平, 等. 配电网三相不平衡度近似计算方法简析[J]. 高压电器, 2019, 55(12): 211-214.
- WU J, ZHU J, SHEN H P, et al. Analysis of the three-phase voltage unbalance computing formula used for distribution network[J]. High Voltage Apparatus, 2019, 55(12): 211-214.
- [16] HUANG CH W, FENG Y B, ZHANG Y B, et al. A GBDT algorithm based approach to power equipment defect data mining and analysis[C]. 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications(ICMNWC), 2022: 1-5.
- [17] 李江腾, 王非. 基于知识嵌入和 DNN 的工商业用户异常用电检测[J]. 电力工程技术, 2020, 39(3): 158-165.
- LI J T, WANG F. Non-technical loss detection based on energy measurement knowledge and deep neural network among industrial and commercial customers[J]. Electric Power Engineering Technology, 2020, 39(3): 158-165.
- [18] DAS S, AGARWAL N, VENUGOPAL D, et al. Taxonomy and survey of interpretable machine learning method[C]. 2020 IEEE Symposium Series on Computational Intelligence(SSCI), 2020: 670-677.
- [19] LUNDBERG S, LEE S I. A unified approach to interpreting model predictions[C]. 31st International

- Conference on Neural Information Processing Systems, 2017.
- [20] 张浩, 谢大智, 胡云晟, 等. 基于半监督学习的网络异常检测研究综述[J]. 信息网络安全, 2024, 24(4): 491-508.
- ZHANG H, XIE D ZH, HU Y SH, et al. A review of network anomaly detection based on semi-supervised learning[J]. Netinfo Security, 2024, 24(4): 491-508.
- [21] 李海霞, 宋丹蕾, 孔佳宁, 等. 传统机器学习模型的超参数优化技术评估[J]. 计算机科学, 2024, 51(8): 242-255.
- LI H X, SONG D L, KONG J N, et al. Evaluation of hyperparameter optimization techniques for traditional machine learning models[J]. Computer Science, 2024, 51(8): 242-255.
- [22] 金自翔, 戴新宇, 陈家骏. 一种基于贪婪算法的 KNN 参数选择策略[J]. 广西师范大学学报(自然科学版), 2008, 26(1): 182-185.
- JIN Z X, DAI X Y, CHEN J J. A strategy for parameter selecting in knn based on greedy search[J]. Journal of Guangxi Normal University(Natural Science Edition), 2008, 26(1): 182-185.
- [23] 吴程昊, 莫路锋. 基于改进狮群算法的支持向量机参数优化方法[J]. 现代电子技术, 2022, 45(14): 79-83.
- WU CH H, MO L F. Method of SVM parameter optimization based on DALSO[J]. Modern Electronics Technique, 2022, 45(14): 79-83.
- [24] YUAN D, HUANG J, YANG X, et al. Improved random forest classification approach based on hybrid clustering selection [C]. 2020 Chinese Automation Congress(CAC), 2020: 1559-1563.
- [25] ZANETTI M, JAMHOUR E, PELLENZ M, et al. A tunable fraud detection system for advanced metering infrastructure using short-lived patterns [J]. IEEE Transactions on Smart Grid, 2019, 10(1): 830-840.

作者简介

王进(通信作者),博士,副教授,主要研究方向为统计机器学习与分布式计算。

E-mail: wangjin@cdu.edu.cn

马智强,硕士研究生,高级工程师,主要研究方向为电能计量、数据采集、数据分析等。

E-mail: mf.512@163.com