

基于机器学习的可再生能源电力系统检修计划优化

方勇,宋涛,郭子强,王健,郭杰

(国网甘肃省电力有限公司兰州供电公司,甘肃 兰州 730000)

摘要:随着可再生能源的不断发展,可再生能源在电网的占比越来越大,为了保证考虑可再生能源接入的电力系统的供电安全,检修工作的最佳协调变得越来越重要。目前检修计划制定工具受运行安全标准约束和电网复杂性的影响,具有可操作性低、模拟意外事故计算量大等问题,为了减轻人工计算负担,提出利用机器学习模型以快速可靠的方式预测应急情况的结果。该方法在兰州某地区设备中进行了测试,涵盖了 10 kV 和 220 kV 的电压等级。通过测试和比较朴素贝叶斯分类器、支持向量机和基于决策树的模型,可知基于决策树的随机森林算法在识别可安全检修时间段的准确率高于 90%,始终优于其他算法。另外,通过实验表明,可再生能源发电的预期增长将影响未来电力系统的可检修性,部分地区非安全检修时间段将增加 20%。

关键词:可再生能源;机器学习;检修计划;可靠性;基于决策树的模型

中图分类号: TM732 **文献标识码:** A **DOI:** 10.19457/j.1001-2095.dqed25090

Optimization of Renewable Energy Power System Maintenance Plan Based on Machine Learning

FANG Yong, SONG Tao, GUO Ziqiang, WANG Jian, GUO Jie

(State Grid Gansu Electric Power Company Lanzhou Power Supply Company, Lanzhou 730000, Gansu, China)

Abstract: With the continuous development of renewable energy sources and the increasing share of renewable energy in the grid, optimal coordination of maintenance work becomes increasingly important in order to ensure the safety of power supply in power systems considering renewable energy access. Current tools for maintenance planning are constrained by operational safety standards and the complexity of the grid, and have problems such as low operability and high computational effort to simulate accidents. To reduce the burden of manual computation, the use of machine learning models was proposed to predict the outcome of emergency situations in a fast and reliable manner. The method was tested in a regional facility in Lanzhou, covering voltage levels of 10 kV and 220 kV. By testing and comparing a plain Bayesian classifier, a support vector machine (SVM) and a decision tree-based model, it was shown that the decision tree-based random forest algorithm is consistently better than other algorithms in identifying safe serviceable time periods with an accuracy rate higher than 90%. In addition, it was shown experimentally that the expected growth in renewable energy generation will affect the future serviceability of the power system, with a 20% increase in non-safe serviceable time periods in some areas.

Key words: renewable energy; machine learning; maintenance plan; reliability; tree-based model

为保证供电安全和质量、以尽可能低的成本促进能源转型,需要对电力系统进行更加细化的安排^[1-2]。然而近年来,可再生电源在电网的占比不断增加,其波动性与不确定性给电网安全运行带来了挑战^[3-5],同时更复杂的发电预测、交通和城市供暖系统等带来的用电需求增加,使得合理进行电网日常调度工作、科学安排检修计划变得越来越困难,可用于安全检修的合适时间段也变得更少^[6]。因此,针对大量可再生能源接入的新

型电力系统,实现可靠性与成本之间的最佳权衡、合理安排检修计划至关重要^[7-8]。

在我国目前的电力系统调度工作中,安排不同电压等级检修计划、寻找最佳停电时间都是基于电网专家的经验 and 知识进行判断,耗费大量的时间、人力和物力。然而由于电网运行的日益复杂化,包括天气、管理需求以及安全规范等因素都需要被考虑,它们存在着复杂的相互依赖关系^[9],如果依旧单纯依靠人为判定,可能对电网安

基金项目:国家电网有限公司科技项目(B3270122000D)

作者简介:方勇(1976—),男,本科,高级工程师,主要研究方向为电网规划建设、高电压技术,Email:275785437@qq.com

全运行造成威胁,甚至导致大规模停电。因此需要创新的检修计划辅助调度工具来制定可靠和灵活的检修计划^[10]。

目前的电网中大量监测设备的部署导致了数据驱动的预测性检修的出现,通过分析电网设备的运行状态来提前发现问题,从而及时进行检修计划的安排,防止意外停电^[11]。文献[12]中基于深度学习辅助检修工作的安排,为调度人员提供关于设备健康使用的相关信息;文献[13]则开发了支持向量机模型,用以预测在极端天气下的电网设备非正常停机;文献[14]将多种不同类型的分类器进行组合,从而改进单一分类器对检修计划的优化。在文献[15]中,基于决策树的模型具有良好的可操作性和准确性,为预测基于设备状态的持续监测故障提供了可能性。

然而,尽管这些基于机器学习的决策工具对于预测设备的非正常故障十分有效,但却无法预测停电部分对整个电网的危害。当配网设备或线路需要检修时,可能使得输电线路过载、无法完全适配电网中的停电调度。此时需要提供转供电计划或者将负荷卸去以实现潮流平衡。鉴于此,可以将机器学习工具作为补充来完善当前的预测性检修策略,该工具能够识别电力系统中检修停电对整个电网带来的影响,确定在不危及电力系统安全的情况下进行检修的时间段。针对类似问题,学术界已进行了一定研究。文献[16]中构建了一种用于预测电网设备检修期间的电力系统运行状况的机器学习工具;文献[17]中介绍了一种计及多类型能源微电网预测检修的优化公式,实现了检修与电网运行的最优控制。但上述模型都不是在满足 $N-1$ 安全校核(即当任意一条线路断开后不会引起系统其它线路出现过负荷或系统解列)运行的标准下进行的。

因此,本文开发了一种综合考虑计划内与计划外检修计划的机器学习结果预测方法,涵盖兰州某地区各电压等级,总共模拟约7 000种计划检修停电与计划外突发状况的组合,以准确描述一年中某时刻的系统可靠性。结果表明,随机森林算法能够实现高准确性地识别安全检修时间,有效辅助检修计划决策,大大减少电网检修工作的成本,提高工作效率。

1 基于应急情况的机器学习模型

安排检修工作时,电网调度负责人会考虑计

划外可能导致电网安全事故的突发情况,预测性地调整电网拓扑结构,使得电力系统符合 $N-2$ 安全校核的运行标准,从而确保电网运行状态处于安全限制范围之内。这类工作通常是精确到检修计划具体小时,通过执行应急预案来分析评估每个小时步长的所有可能停电事故。

由于应急研究耗时且繁琐,为了简化应急分析,突出研究重点,本文开发了一种监督学习模型,利用现有数据来学习 ID 中的预测因子 x_t^{orig} (即 t 时刻的负荷、发电量和天气条件等)与设备 c 的检修可行性 $y_{c,t}$ 之间的关系。然后,经过多次训练后的模型可用于可靠判断任一新的系统状态是否可以安全地执行检修任务。

图1总结了机器学习模型的全局程序思路。本文中考虑的设备包括输电线路和变压器,首先构建数据库,将可用的输入特性 $x_t^{\text{orig}} \in \mathbf{R}^{|D|}$ 与所有时刻 $t \in T$ 的所有设备 $c \in C$ 检修可行性 $y_{c,t} = \{0,1\}$ 一一对应。由于每个电网设备 c 受到不同参数的影响,因此在数据库构建后,对输入特性进行选择,使得 ID_c 中保留最相关的预测因素 $x_c \subset x^{\text{orig}}$ 。最后,基于所选择的输入特性,为每个设备 c 训练一个专门的机器学习模型 f_{θ_c} 。

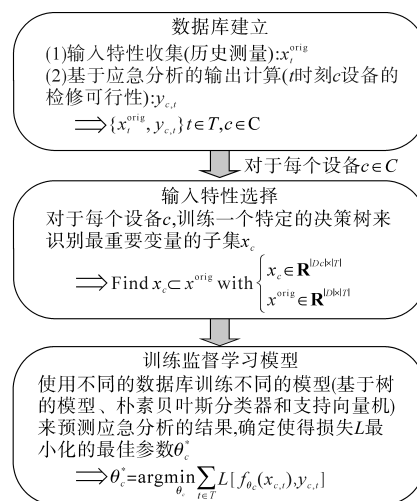


图1 机器学习模型的全局程序思路

Fig.1 Global program ideas for machine learning agents

为确保该方法适用于电网,模型还需要遵守3个重要的限制条件:1)需要保证其易于使用:模型需要在没有复杂的数据预处理的情况下进行训练。2)模型需要具备较强的鲁棒性:能够对参数变化做出响应,无需依靠专家知识与经验。3)所得结果呈现形式简单:能够识别重要的输出特征。本文基于上述3个限制条件,采用不同类型的机器学习模型与基于决策树的随机森林模型

进行比较。

1.1 数据库建立

图2为构建用于训练机器学习模型的数据库流程图。数据库 $x^{\text{orig}} \in \mathbf{R}^{|D| \times |T|}$ 由 $t \in T$ 的每个小时时间段所测得的变量 $d \in D$ 组成。其中 $|D|$ 包括全局数据与局部数据。

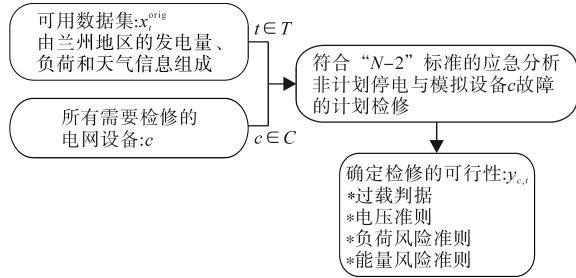


图2 创建用于训练机器学习模型的数据库

Fig.2 Creation of the database used for training the machine learning models

在数据库中可获得地区的总负荷量、依靠不同技术的综合发电量以及天气信息。总负荷量包括从非高峰期到高峰期的总负荷量变化;综合发电量组合为核电、天然气、海上和陆上风电、光伏、沼气、煤炭以及其他;而天气信息则是由地区的温度、云量和风速构成。程序中,本地变量是指传输系统不同节点的发电量和耗电量测量,节点包括大型工业用电、发电厂与本地配电网接口。对于兰州某地区而言,输电网上有1 000多个节点,显著增加了计算的维度。通常,对于历史实现的数据集 $\{x_t^{\text{orig}}\}_{t \in T}$ 可用于训练机器学习模型,然而这些输入场景数据集也可以通过市场模拟生成,建立范围广泛且历史数据库中不存在的原始电网条件。根据模拟预期的市场条件,可构建推广至未来的模型,实现对未来市场的预测。本文中的模拟场景与获取的历史信息中不存在缺失数据,因此对于原始数据无需进行数据整合。

作为对数据库的补充,算法还列出了所有需要检修的不同电网设备。对于每个设备 $c \in C$ 的目标是计算检修数据库每个时刻 $t \in T$ 的可行性,同时满足执行“N-2”安全校核标准的应急分析,模拟所有与计划检修同时发生的非计划停电。换言之,当可检修设备 c 满足以下4个条件时,则设备可检修:

1)过载判据:剩余电网中的任何设备都不能发生过载现象。因为这可能会导致电网中发生电压崩溃或者负荷损失的连带效应和负面影响。

2)电压准则:电网的每个节点应保持在规定

的电能质量标准内(即额定值的90%~105%)。

3)负荷风险准则:在突发事件发生后,电网用户可能损失的总功率保持在规定阈值以下。

4)能量风险准则:向电网用户供应的总能量不得超出规定阈值。

当不符合上述标准的情况发生时,设备则不能进行检修,即 $y_{c,t} = 0$ 。同时,对于上述多标准方法还嵌套可靠性与经济性的指标来定义检修的可行性^[18]。

1.2 相关输入变量的选择

由于数据集包含有许多不同变量,并且每个电网设备都会受到一组不同变量的影响,所以对于每个单独的设备应对其进行特定的变量选择,从而减少输入数据的维度以及初始数据的损失。通过这种特定性选择,能够降低模型的复杂性,减少训练时间,有效避免过拟合的风险。

为此可采用不同的方法对相关输入变量进行选择。主成分分析(principal component analysis, PCA)和线性判别分析(linear discriminant analysis, LDA)的主要思想是将高维数的输入数据线性映射到一个新的低维数空间,此空间包含有设备的所有相关信息;核主成分分析(kernel PCA, KPCA)则可实现数据的非线性降维,用于处理线性不可分的数据集;自动编码器给出专门用于学习数据的缩小空间表示的神经网络,提供了另一种有效的非线性框架。然而,上述方法的数据更新和简化都没有真实的物理意义,既无法解释说明也不能与原始数据进行相关的特征比较,方法得到的数据与所选择的待评估设备的实际特征没有直接的相关性。鉴于此,本文采用决策树的方式解决上述变量选择的困难,它能够对来自原始数据集的重要变量进行识别,收集有价值的属性而不需要任何特征缩放,并且其本身能够生成多变量的非线性依赖,不需要依赖于特定的外部假设。决策树算法能够有效地选择设备 c 的最相关子集 D_c ,同时在后续分析可知同样也可将决策树用于分类以预测维修的可行性。

决策树是一种监督学习方法,能够将整个 $|D|$ 维特征空间划分成更小的子空间。以图3所示的带有两个变量的决策树为例,决策树是通过分层的多阶段过程来训练的。在每个阶段中,创建新的机会节点 m ,采用算法选择来提供数据之间最佳分离的变量,实现不同类别的最佳区分。基于阈值 M 将变量分为两个子空间后分别记为 m_1 和

m_c 。程序一直执行到进一步划分的输入空间不再给分类精度带来价值为止,最后节点显示为最终分类结果。根节点和节点2为机会节点,节点3、节点4和节点5为端节点。

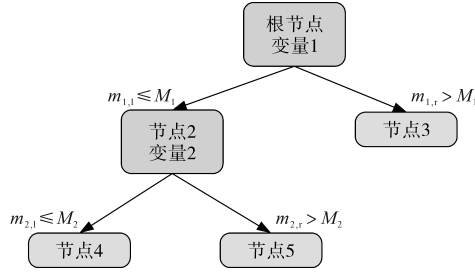


图3 带有两个变量的决策树示例

Fig.3 Illustrative example of decision tree with two variables

在每个阶段中,由要划分的最佳特征 m 及其相应的阈值 M 组成的最佳模型参数是基于不纯度函数构建的,本文中采用基尼不纯度如下所示:

$$Gini(m) = 2p_m(1 - p_m) \quad (1)$$

式中: p_m 为设备检修可能被归类在节点 m 的概率。

从式(1)可知,基尼指数在 0~0.5 之间变化。当节点 m 与关联该节点的所有样本都属于一个类别时,基尼指数为 0;而当节点的所有样本处于可检修与不可检修两个等级之间平均分配时,基尼指数等于 0.5。

一旦对应于设备 c 的决策树完成训练,就可根据不同变量在树中的位置来确定 D_c 的最重要变量。通过这种方式,变量的重要性被计算为节点不纯度的减少乘以到达该节点的概率。由于单个变量可以在树的不同节点上使用,因此它的整体重要性是其单个贡献的总和。然而,这种重要性变量排名并没有提供关于临界值的信息,在该临界值下,变量对于分类没有真正的作用。为了以通用的方式识别适用于每个设备的临界值,将一个随机变量添加到预测因子 x^{orig} 中,并在大小为 $\mathbf{R}^{(|D|+1) \times |T|}$ 的扩充数据集上训练树。由于随机变量对于设备可检修性没有影响,因此可以从数据集中删除所有得分较低的变量。最后为了确保结果的稳定性,重复 10 次操作并对每个变量的得分进行平均。

1.3 分类模型

在对相关变量选择完成后,需要开发一个快速可靠的工具以实现在规定电网条件下预测设备的可维护性,该问题是一个具有二进制结果的分类任务,即当设备可安全检修时 $y_{c,d} = 1$;当检修设备会导致剩余系统出现故障时 $y_{c,d} = 0$ 。工

具目标是优化设备模型 f_{θ_c} 的参数 θ_c ,根据给定的输入 $x_{c,d}$ 准确预测输出 $y_{c,d}$:

$$\theta_c^* = \arg \min_{\theta_c} \sum_{t \in T} \{L[f_{\theta_c}(x_{c,t}), y_{c,t}]\} \quad (2)$$

式中: L 为在学习阶段用于减少预测 $\hat{y}_{c,d}$ 与实际 $y_{c,d}$ 输出之间的分类误差的损失函数。

训练结束后,得到模型的最优参数 θ_c^* 可用于预测当设备检修条件下的应急分析结果 $\hat{y}_{c,d}$ 。针对不同的机器学习模型,测试具体过程如图 4 所示。

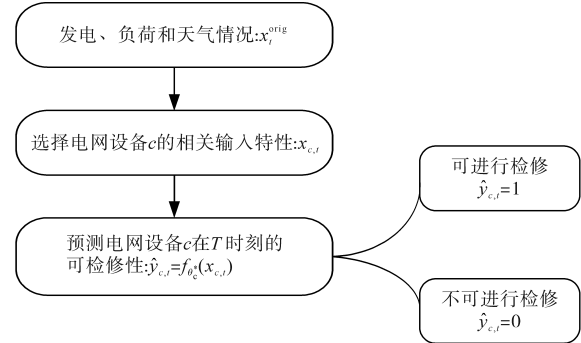


图4 机器学习模型的全局程序思路

Fig.4 Global program ideas for machine learning agents

1.3.1 决策树

如前所述,决策树的训练阶段采用的是贪婪搜索,其中节点 m 处的最优模型参数 $\theta_{c,m}^*$ 是特征 m 的最优分割。在模型训练的过程中,要通过控制树的深度等方式来避免过拟合问题,从而确保对未处理数据的最佳优化。训练后,将决策树算法与朴素贝叶斯分类器和支持向量机(support vector machine, SVM)所得结果进行比较。

1.3.2 朴素贝叶斯分类器

朴素贝叶斯分类器依赖于 $|D_c|$ 的输入 $x_{c,d} = (x_{1,d}, \dots, x_{|D_c|,d})$ 完全独立,对于预测结果 $y_{c,d} = \{0, 1\}$ 的贡献完全相等。因此贝叶斯定理可以表示为^[19]

$$P(y_{c,d} | x_{1,d}, \dots, x_{|D_c|,d}) = \frac{P(y_{c,d}) \prod_{i=1}^{|D_c|} P(x_{i,d} | y_{c,d})}{P(x_{1,d}, \dots, x_{|D_c|,d})} \quad (3)$$

由于 $P(y_{c,d} | x_{1,d}, \dots, x_{|D_c|,d})$ 对于给定的数据库是常数,可以使用以下分类规则:

$$\hat{y}_{c,d} = \arg \max_{y_{c,d}} P(y_{c,d}) \prod_{i=1}^{|D_c|} P(x_{i,d} | y_{c,d}) \quad (4)$$

其中, $P(y_{c,d})$ 和 $P(x_{i,d} | y_{c,d})$ 易使用最大后验估计进行计算,此方法下 $P(y_{c,d} = 0)$ 是训练集中 $y_{c,d} = 0$ 的相对频率,因此 $(y_{c,d} = 0) + P(y_{c,d} = 0) = 1$ 。

1.3.3 支持向量机

支持向量机旨在 $|D_c|$ 维空间中找到一个 $|D_c|-1$ 维超平面,该超平面对数据点 $y_{c,d} = 0$ 和 $y_{c,d} = 1$ 进

行明确分类。在所有可能的超平面中,目标是找到一个导致两类点之间的距离最大的超平面,使得后续结果能够更加准确地分类,优化模型如下:

$$\begin{cases} \max_{\alpha} = \sum_{t=1}^{|T|} \alpha_t - \frac{1}{2} \sum_{t=1}^{|T|} \sum_{u=1}^{|T|} \alpha_t \alpha_u y_{c,t} y_{c,u} K(\mathbf{x}_{c,t}, \mathbf{x}_{c,u}) \\ \text{s.t. } 0 < \alpha_t < C \\ \sum_{t=1}^{|T|} \alpha_t y_{c,t} = 0 \end{cases} \quad (5)$$

式中: C 为用于惩罚训练误差的常数; K 为用于将输入特征映射到高维空间中的核函数,能够有效处理复杂分类问题。

支持向量机中,线性核函数 $K(\mathbf{x}_{c,t}, \mathbf{x}_{c,u}) = \mathbf{x}_{c,t}^T \mathbf{x}_{c,u}$ 产生了良好的效果。通过求解式(5),可获得正则化参数 α ,该参数在两类样本之间提供了更宽的裕度。

1.3.4 初步选型

通过对 3 种模型进行测试,预测兰州某地 110 kV 线路在不同条件下检修的可行性。本文获取了线路一整年的工作数据,并将其细分为训练集和测试集,训练集由 70% 数据组成,测试集则由剩下 30% 组成。将成功率定义为测试集 2 621 个点上准确分类为可检修和不可检修样本的数量,因此不同模型的准确性可采用成功率 r 来进行对比。3 种模型在训练之前的不同之处在于,决策树和朴素贝叶斯分类器不需要进行数据预处理,而 SVM 则需要在训练模型之前通过下式所示的 Min-Max 技术将数据归一化在 $[0, 1]$ 之间。

$$\mathbf{x}_{d,t}^{\text{orig}} = \frac{\mathbf{x}_{d,t}^{\text{orig}} - \min(\mathbf{x}_d^{\text{orig}})}{\max(\mathbf{x}_d^{\text{orig}}) - \min(\mathbf{x}_d^{\text{orig}})} \quad \forall d \in D \quad (6)$$

式中: $\mathbf{x}_d^{\text{orig}}$ 为与变量 $d \in D$ 相关联的值的向量。

为了不同分类器之间对比的公平,通过最大范围的搜索找到每种模型超参数的最佳组合,从而最大限度地提高泛化能力和分类精度^[20]。每种技术训练了具有不同超参数值的模型,并将模型根据交叉验证程序实现的准确性来选择最佳模型。结果总结如表 1 所示。

表 1 不同分类器的结果矩阵

Tab.1 The resulting matrix of the different classifiers

	朴素贝叶斯分类器		支持向量机		决策树	
	$\hat{y}_0$	$\hat{y}_1$	$\hat{y}_0$	$\hat{y}_1$	$\hat{y}_0$	$\hat{y}_1$
y_1	112	2 114	18	2 508	16	2 510
y_0	92	3	35	60	70	25
r	95.6%		96.4%		98.4%	

从表 1 可知,决策树在成功率方面优于其他

两种分类器,得分高达 98.4%。决策树预测检修在 16 h 内无法执行,在这 16 h 内,系统本可以安全完成并将安全检修时间延长至 25 h,而之所以会出现无法检修的情况,是因为若进行 25 h 检修,则实际上传动系统会出现可靠性问题,从而导致事故发生。SVM 模型偏向于预测安全检修。朴素贝叶斯分类器的整体效率偏低,但其提供了更为安全的建议,只出现了 3 个假性可检修情况。

通过上述结果可知,由于决策树的高准确性以及对检修工作人员的有用属性,选择其作为分类工具是正确进行研究的关键。它被直观地理解并且可以通过任何类型的输入特征来提供,而不需要将数据进行规范化,省去了数据库进行预处理的过程。

1.3.5 分类模型改进

为了进一步提高基本决策树算法的性能而不改变其易用性和可解释性,对分类模型进行了不同的改进。其思想是结合多个决策树以减少结果模型的可变性。这种基于决策树的模型具有很高的预测精度,即使在具有高度相关特征的高维问题中也具有良好的效果。Bagging 算法和 Boosting 算法作为两种改进算法可较好地适应检修工作的分类。Bagging 算法以随机森林为代表,通过创建单步 N^{RF} 独立决策树,每一棵树建立在特征的随机子集上且每棵树的每次分割都是基于剩余数据集的随机子样本构建的,之后采用平均 N^{RF} 独立决策树的结果来进行预测。Boosting 算法则以梯度增强决策树和极端梯度增强决策树为代表,通过依次创建新的模型来预测前一阶段获得的全局模型的残差。

2 案例分析

本文将所提出算法用于兰州某地区获得的实际数据库中,3 种分类工具使用 R 语言编程实现并在 Python 中搭建潮流网络实现应急分析,进行不同电网设备停电下电力系统状态的潮流计算。

2.1 评估指标

对于电力系统来说,错误预测安全停电的情况应尽可能减少,因此可根据电网的错误分类成本来控制是否真正可进行检修。具体实现则是通过调整阈值 ρ 来判断检修可行性。通常,对于二进制的分类任务,该阈值固定为 $\rho = 0.5$,表示当分类器产生的概率高于 50%,没有意外情况导致不可靠或者经济损失的情况发生,则认为

设备可检修。然而当设置 $\rho > 0.5$ 时可降低假阳性(错误的可检修)的发生率。为了指导决策阈值 ρ 的最佳选择,采用接受者操作特性曲线(receiver operating characteristic, ROC)进行分析。

ROC曲线显示不同阈值 ρ 的真阳性(真正的可检修)率和假阳性率的关系。一个完美的分类器存在一个包含点(0,1)的ROC曲线,对应于没有误报和100%的真阳性。本文中的最佳决策阈值 ρ^* 被定义为导致ROC曲线和最佳点(0,1)之间的最小欧几里得距离的阈值。同时,ROC曲线下的面积也可用于比较不同分类模型的准确性,当分类器的一条线性ROC曲线穿过模型无法区分类别的点(0.5,0.5)时,此时ROC下的面积对应值为0.5。而当曲线穿过的是最佳点(0,1)时,面积对应值为1。面积对应值越大,则模型准确性越高。

2.2 基于决策树的模型性能

为评估基于决策树的模型在预测应急分析结果方面的性能,判断能否在不危及电网运行安全的情况下对某一设备进行检修,将兰州某地区电网一年的运行数据分为70%的训练集和30%的测试集,并针对两种特性设备进行分析。

2.2.1 设备1

首先,选取兰州某地区的一条主要高压输电线路作为设备1,通过在整个数据库上训练决策树来选择最相关的解释变量,预测不同条件下检修的可行性。由于输入数据变量具有高维性,因此训练后的模型无法以图形方式表示,但通过1.2节的方法可见从训练模型中量化变量的重要性,结果如图5所示。

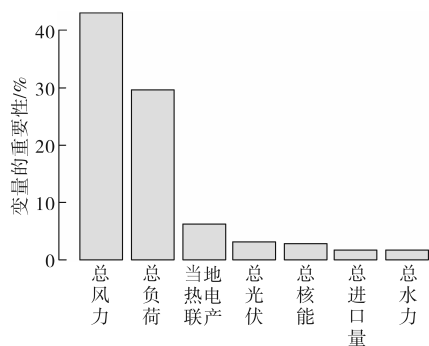


图5 影响设备1检修可行性的重要变量排序

Fig.5 Ranking of important variables affecting the feasibility of equipment 1 overhaul

从图5可知设备1的可检修性受到兰州某地区总负荷、风能以及光伏发电总量的强烈影响。这是由于设备1选取的是输电系统的主干架空线,它主要受分布式电源功率交换的影响是与实

际情况是一致的。从结果可知,由于无法在长期内可靠地预测重要的影响变量,因此定义所研究线路的检修计划并不能简单地实现。

通过量化设备1在训练和测试数据上的性能来分析决策树的准确性,采用ROC曲线分析结果如图6所示。

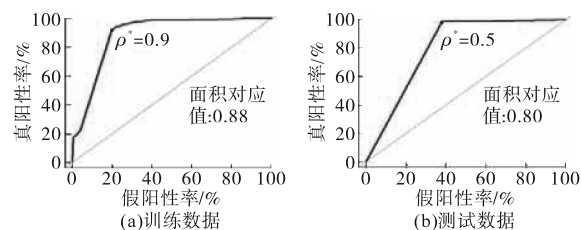


图6 决策树的性能

Fig.6 Performance of decision tree

根据已有文献研究表明,决策树具有较高的方差,容易导致过度拟合,使得训练和测试数据之间的准确性明显下降。如图6所示,当新获取的数据被送入训练模型时,面积对应值从训练中的0.88下降到0.8。因此,本文将决策树仅用于识别训练期间最重要的变量并量化变量的重要性,而实际预测是使用基于决策树的高级模型执行的。

基于所选变量,可以训练随机森林,将决策树的数量作为一个重要的超参数,采用平均值来减少模型的方差。当树的数量增加时,生成的模型需要更多的时间来训练,因此需要找到准确性与模拟时间之间的最佳权衡。在图7中,通过向随机森林中的每棵树提供学习过程中未使用的样本并采用平均分类误差来描述随机森林的袋外误差。

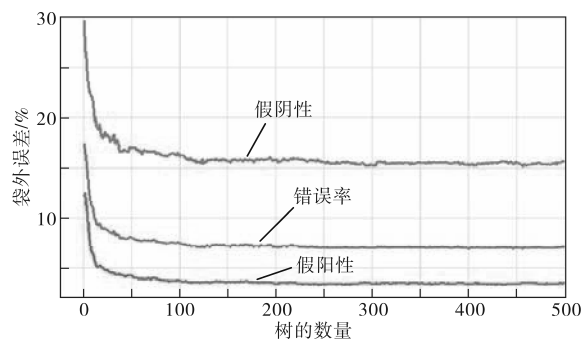


图7 袋外误差作为随机森林中树木数量的函数

Fig.7 Out-of-bag error as a function of the number of trees in the random forest

从图7中可以看出,最佳权衡方案是在随机森林中使用大约100棵决策树。当数量大于100棵时,模型的性能增益微乎其微,但模拟时间却大大增加。图中假阳性和假阴性的数量与决策

阈值的默认值 $\rho=0.5$ 相关。

为了实现最大化性能,同时限制过拟合,优化随机森林模型的其他超参数也同样重要。本文采用75%的特征空间训练单独的树以形成随机森林,并且通过至少5个样本来分割1个内部节点,但是并不对每棵树的深度进行限制,即输入空间中允许分割的最大数量是可以任意确定的。对于梯度增强决策树来说,为达到精度和计算时间之间的良好平衡,将Boosting阶段的数量固定为100。同时,另一个需要优化的超参数是学习率,它是对每棵树的贡献进行加权,从而不断更新模型,本文固定学习率为0.1。当超参数固定后,随机森林与梯度增强决策树模型同时训练,并在测试集上进行评估,结果如图8所示。

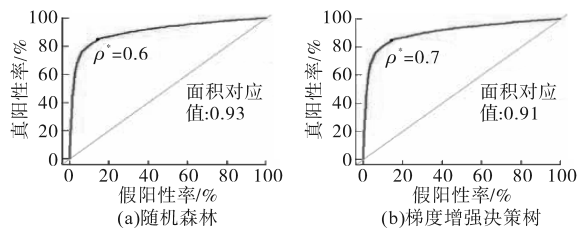


图8 不同模型训练的ROC曲线

Fig.8 ROC curves trained by different models

从图8中可以看到,两个基于决策树的模型都表现出较高的准确性且最佳决策阈值都比传统算法更加保守,其中随机森林略优于梯度增强决策树。两种模型的ROC曲线都非常接近代表完美分类器的理论最佳点(0,1),而由于随机森林对超参数的变化具有更好的鲁棒性且计算时间比梯度增强决策树快了2倍,因此后续研究仅采用随机森林算法。

最后,为了量化变量选择程序,将图5所示的解决方案与其他两个基准进行比较。第一种情况下只选择两个最重要的变量,而第二种情况考虑所有的输入变量。将通过不同方法选择的变量分别送入随机森林分类工具,得到最终结果。具有两个变量的模型面积对应值为0.87,相对于使用最佳变量数的模型面积对应值损失了0.06,即准确性下降。而具有所有变量的随机森林模型面积对应值达到了0.94,虽然准确性很高,但处理高维输入空间会使得训练过程显著复杂化。

2.2.2 设备2

将随机森林法用于电网不同区域的电力变压器,专用决策树用于确定最重要的输入特征,如图9所示。

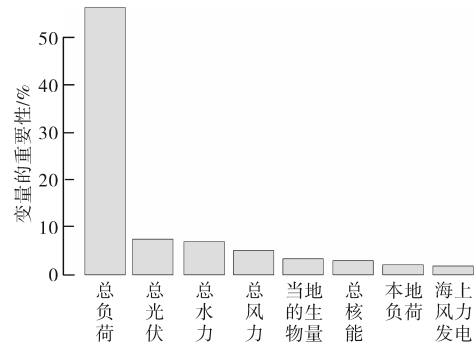


图9 影响设备2检修可行性的重要变量排序

Fig.9 Ranking of important variables affecting the feasibility of equipment 2 overhaul

该设备的可重点性主要受负载消耗的影响。光伏、水电以及生物质能的影响有限。利用所选变量训练随机森林得到的设备2的ROC曲线如图10所示。

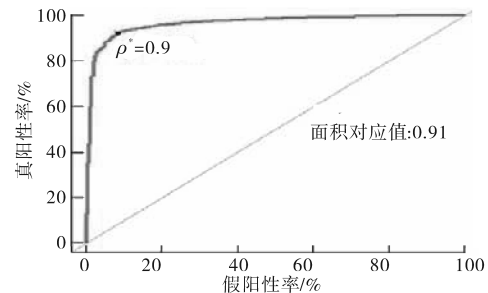


图10 设备2的随机森林ROC曲线

Fig.10 ROC curve of the random forest for the equipment 2

从图10可知,随机森林模型的准确性非常高,面积对应值达到了0.91。最佳决策阈值是针对 $\rho=0.9$ 的保守值获得的,然而,这个阈值仍会导致大约10%的误判。可通过ROC曲线来选择更加保守的截止 ρ ,但同时需要注意系统中仍应该存在能够保留检修工作的中断窗口。针对不同设备的大量测试,随机森林算法取得了类似的结果,因此这些模型可以继承至规划工具中,同时在得到所有结果后,可快速预测全局的可检修性。

2.3 模型开发

本文开发的机器学习模型可服务于两个主要目的。首先,从电网运行的角度来看,该模型可用于直接向检修计划工具告知电网设备可安全停运的时间段,避免在决策过程中模拟所有的意外停机事件。电网调度部门可根据获得的新信息动态调整停电时间表。其次,鉴于电网输电线路和变压器等设备众多,了解所有设备是否可检修能够帮助电网制定合理的检修计划。

为了对系统的可检修性进行全局量化,开发了两个互补的指标。第一个得出了安全时间段

的普遍性。在此期间,计划停运有望最大限度地减少可靠性和经济因素。第二个则描述了准确预测合适的停运时间段的复杂性。这些指标是根据兰州某地区三个不同区域计算的,用以评估区域运行状态。

2.3.1 临界性和不可预测性指标

首先,通过定义临界分数来描述为安排安全停机找到足够时间的难度。该指标定义为一年内设备的检修工作在总工作日中被定义为至少1 h不安全的工作日数。当该指标为0代表没有合适的时间安排检修,为1则代表设备可安全停运而不影响电网安全,数值在0到1之间变化。

由于确定设备能否安全停机可能受到复杂的预测变量显著影响,因此需要根据预测变量的重要性定义一个评分来量化检修工作的不可预测性,从而作为临界分数的补充。具体计算是将不可预测变量的总贡献除以所有输入变量的贡献,0分表示设备是否可检修仅取决于容易预测的变量,1分表示设备的运行状态受到更复杂的预测变量的影响。

2.3.2 特定区域的全局可检修性

选取兰州某地区三个不同区域进行指标计算,所有设备的得分如图11所示。不同的符号表示位于不同区域的设备。为了便于解释,将图分为简单、挑战以及困难三个部分。

从图11中可知,大部分的设备位于图的左下角,表示三个区域的大部分设备具有良好的可检修性,同时可观察到设备主要受可预测变量的影响。另外,由于全年可用的安全时段较多,因此大部分检修工作能在不危及电网安全的情况下进行,表明兰州电力系统具有良好的稳定性以及抗突发事件的能力。

对于远离最佳点(0,0)的设备,具有较高的不可预测性和临界性指标值,需对此类设备进行关注。本文通过调查未来几年区域可检修性的演变来解决此类问题。具体是预测未来气候年的代表性情况(如有大量风或者温度较高的年份)、未来负荷行为以及发电组合,然后将这些场景作为训练基于决策树的模型的输入,从而预测所研究年份的每个小时的每个设备的可使用性。程序针对2024年、2025年和2030年的三个区域电网设备进行实现。结果如图12所示,对于此训练,随机森林能够在15 min内完成。

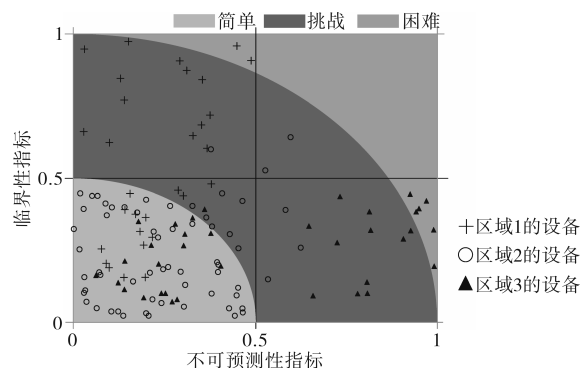


图11 三个不同区域内所有设备的临界性与不可预测性得分

Fig.11 Criticality and unpredictability scores for all devices in three different areas

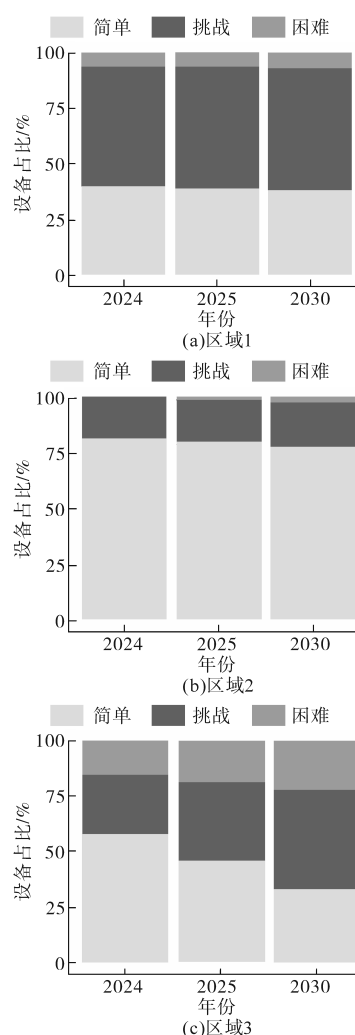


图12 基于预测输入情景的区域可检修性演变

Fig.12 Evolution of regional maintainability based on predicted input scenarios

图12中预测的3个研究区域的可检修性将遵循不同的趋势。从图11中观察可知区域3有许多设备受到可再生能源发电的影响,结合图12分析,此类设备将随着能源的预计增加而增加检修的难度。而区域1和区域2设备的未来可检修

性可保持稳定,仅在2030年略有下降。因此,可将75%的区域2设备安排在2030年。

3 结论

本文提出了一种改进现代电力系统安全约束的辅助检修计划新方法,该方法适用于兰州某地区电网的不同设备。通过测试不同的机器学习技术预测特定电网条件下检修的可行性,为电网检修规划提供有效帮助。

首先,在输入变量选择过程中,为了避免过拟合的风险,在确定最重要的输入变量的同时,增加了输入选择过程进行优化。

其次,对比了三种模型预测情况,其中随机森林的预测精度最高,成功率在90%以上,当进一步减少输入空间则会使得预测精度下降。

最后,对未来兰州某地区3个区域进行模拟预测,表明可再生能源发电的普及可能对系统的可检修性产生影响,甚至会造成安全检修时间段减少20%的情况。

本研究开发的随机森林模型将机器学习与电力系统检修工作结合,辅助检修计划工具能够较好地提高电网检修效率,降低成本。

参考文献

- [1] 卓振宇,张宁,谢小荣,等.高比例可再生能源电力系统关键技术及发展挑战[J].电力系统自动化,2021,45(9):171-191.
ZHUO Zhenyu, ZHANG Ning, XIE Xiaorong, et al. Key technologies and developing challenges of power system with high proportion of renewable energy[J]. Automation of Electric Power Systems, 2021, 45(9): 171-191.
- [2] 刘天琪,张琪,何川.考虑气电联合需求响应的气电综合能源配网系统协调优化运行[J].中国电机工程学报,2021,41(5):1664-1677.
LIU Tianqi, ZHANG Qi, HE Chuan. Coordinated optimal operation of electricity and natural gas distribution system considering integrated electricity-gas demand response[J]. Proceedings of the CSEE, 2021, 41(5): 1664-1676.
- [3] 盛万兴,吴鸣,季宇,等.分布式可再生能源发电集群并网消纳关键技术及工程实践[J].中国电机工程学报,2019,39(8):2175-2186.
SHENG Wanxing, WU Ming, JI Yu, et al. Key techniques and engineering practice of distributed renewable generation clusters integration[J]. Proceedings of the CSEE, 2019, 39(8): 2175-2186.
- [4] 程耀华,张宁,王佳明,等.面向高比例可再生能源并网的输电规划方案综合评价[J].电力系统自动化,2019,43(3):33-42,57.
CHENG Yaohua, ZHANG Ning, WANG Jiaming, et al. Comprehensive evaluation of transmission network planning for integration of high-penetration renewable energy[J]. Automation of Electric Power Systems, 2019, 43(3): 33-42, 57.
- [5] 施贵荣,孙荣富,徐海翔,等.大规模集群可再生能源有功分层协调控制策略[J].电网技术,2018,42(7):2160-2167.
SHI Guirong, SUN Rongfu, XU Haixiang, et al. Active power stratification coordination control strategy for large-scale cluster of renewable energy[J]. Power System Technology, 2018, 42(7): 2160-2167.
- [6] BERTLING L, ALLAN R, ERIKSSON R. A reliability-centered asset maintenance method for assessing the impact of maintenance in power distribution systems[J]. IEEE Transactions on Power Systems, 2005, 20(1): 75-82.
- [7] 杜勉,易俊,郭剑波,等.以可靠性为中心的维修策略综述及其在海上风电场运维中的应用探讨[J].电网技术,2017,41(7):2247-2254.
DU Mian, YI Jun, GUO Jianbo, et al. Review on reliability centered maintenance strategy and applications to offshore wind farm operation and maintenance[J]. Power System Technology, 2017, 41(7): 2247-2254.
- [8] 赵登福,段小峰,张磊.考虑设备状态和系统风险的输电设备检修计划[J].西安交通大学学报,2012,46(3):84-89.
ZHAO Dengfu, DUAN Xiaofeng, ZHANG Lei. Maintenance scheduling of power transmission equipment maintenance scheduling of power transmission equipment[J]. Journal of Xi'an Jiaotong University, 2012, 46(3): 84-89.
- [9] JEAN F T, MORSTYN T, BOTTIEAU J, et al. Capturing spatio-temporal dependencies in the probabilistic forecasting of distribution locational marginal prices[J]. IEEE Transactions on Smart Grid, 2021, 12(3): 2663-2674.
- [10] 张智刚,夏清.智能电网调度发电计划体系架构及关键技术[J].电网技术,2009,33(20):1-8.
ZHANG Zhigang, XIA Qing. Architecture and key technologies for generation scheduling of smart grid[J]. Power System Technology, 2009, 33(20): 1-8.
- [11] NATHALIE Cauchi, KAREL Macek, ALESSANDRO Abate. Model-based predictive maintenance in building automation systems with user discomfort[J]. Energy, 2017, 138: 306-315.
- [12] 孙国强,沈培锋,赵扬,等.融合知识库和深度学习的电网监控告警事件智能识别[J].电力自动化设备,2020,40(4):40-47.
SUN Guoqiang, SHEN Peifeng, ZHAO Yang, et al. Intelligent recognition of power grid monitoring alarm event combining knowledge base and deep learning[J]. Electric Power Automation Equipment, 2020, 40(4): 40-47.
- [13] 叶林,刘鹏.基于经验模态分解和支持向量机的短期风电功率组合预测模型[J].中国电机工程学报,2011,31(31):102-108.
YE Lin, LIU Peng. Combined model based on EMD-SVM for short-term wind power prediction[J]. Proceedings of the CSEE,

- 2011,31(31):102-108.
- [14] SUSTO G A, SCHIRRU A, PAMPURI S, et al. Machine learning for predictive maintenance: a multiple classifier approach[J]. IEEE Transactions on Industrial Informatics, 2015, 11(3): 812-820.
- [15] 李杰其, 胡良兵. 基于机器学习的设备预测性维护方法综述[J]. 计算机工程与应用, 2020, 56(21): 11-19.
LI Jieqi, HU Liangbing. Review of machine learning for predictive maintenance[J]. Computer Engineering and Applications, 2020, 56(21): 11-19.
- [16] 张友强, 钟加勇, 魏甦, 等. 基于贝叶斯网络的智能变电站继电保护设备状态检修方法[J]. 科技通报, 2019, 35(10): 101-104.
ZHANG Youqiang, ZHONG Jiayong, WEI Su, et al. A state maintenance method for Relaying protection equipment intelligent substation based on bayesian network[J]. Bulletin of Science and Technology, 2019, 35(10): 101-104.
- [17] GARGARI M Z, HAGH Z, ZADEH S G, et al. Preventive maintenance scheduling of multi energy microgrid to enhance the resiliency of system[J]. Energy, 2021, 221: 119782.
- [18] 刘文颖, 谢昶, 文晶, 等. 基于小生境多目标粒子群算法的输电网检修计划优化[J]. 中国电机工程学报, 2013, 33(4): 141-148, 19.
LIU Wenying, XIE Chang, WEN Jing, et al. Optimization of transmission network maintenance scheduling based on niche multi-objective particle swarm algorithm[J]. Proceedings of the CSEE, 2013, 33(4): 141-148, 19.
- [19] 吴欣, 郭创新, 曹一家. 基于贝叶斯网络及信息时序属性的电力系统故障诊断方法[J]. 中国电机工程学报, 2005, 25(13): 14-18.
WU Xin, GUO Chuangxin, CAO Yijia. A new fault diagnosis approach of power system based on bayesian network and temporal order information[J]. Proceedings of the CSEE, 2005, 25(13): 14-18.
- [20] BOTTIEAU J, HUBERT L, GRÈVE Z D, et al. Very-short-term probabilistic forecasting for a risk-aware participation in the single price imbalance settlement[J]. IEEE Transactions on Power Systems, 2020, 35(2): 1218-1230.

收稿日期: 2023-04-14

修改稿日期: 2023-06-02

(上接第48页)

- on particle swarm optimization for LCL grid-connected inverter[J]. Electric Power Automation Equipment, 2021, 41(11): 174-182.
- [13] 耿强, 李亮, 周湛清, 等. 双永磁电机系统抗扰动转速同步控制[J]. 中国电机工程学报, 2021, 41(19): 6787-6796.
GENG Qiang, LI Liang, ZHOU Zhanqing, et al. Speed synchronization control of disturbance rejection of dual-PMSM system[J]. Proceedings of the CSEE, 2021, 41(19): 6787-6796.
- [14] 朱焜秋, 顾志伟. 基于模糊神经网络逆系统的五自由度无轴承永磁同步电机自抗扰控制[J]. 电机与控制学报, 2021, 25(2): 72-81.
ZHU Huangqiu, GU Zhiwei. Active disturbance rejection control for 5-degree-of-freedom bearingless permanent magnet synchronous motor based on inverse system using the fuzzy neural network[J]. Electric Machines and Control, 2021, 25(2): 72-81.
- [15] 朱进权, 葛琼璇, 王晓新, 等. 基于自抗扰和负载功率前馈的高速磁悬浮系统PWM整流器控制策略[J]. 电工技术学报, 2021, 36(2): 320-329.
ZHU Jinquan, GE Qiongxuan, WANG Xiaoxin, et al. Control strategy for PWM rectifier of high-speed maglev based on active disturbance rejection control and load power feed-forward[J]. Transactions of China Electrotechnical Society, 2021, 36(2): 320-329.
- [16] 马幼捷, 陶珑, 周雪松, 等. 结合自抗扰的风电系统电压环模糊自适应控制[J]. 太阳能学报, 2020, 41(12): 330-337.
MA Youjie, TAO Long, ZHOU Xuesong, et al. Fuzzy adaptive control of voltage loop in wind power system combined with linear active disturbance rejection control[J]. Acta Energetica Sinica, 2020, 41(12): 330-337.
- [17] 马明, 廖鹏, 蔡雨希, 等. LCL并网逆变器的自抗扰控制策略[J]. 高电压技术, 2021, 47(6): 2223-2231.
MA Ming, LIAO Peng, CAI Yuxi, et al. Active disturbance rejection control strategy of LCL grid-connected inverter[J]. High Voltage Engineering, 2021, 47(6): 2223-2231.

收稿日期: 2023-06-02

修改稿日期: 2023-07-07