

面向无人机巡检场景基于 QRGBDT 的架空输电线路弧垂区间预测

武云发, 林文帅, 段 恒, 谢家豪

广东工业大学自动化学院, 广东 广州

收稿日期: 2023 年 4 月 17 日; 录用日期: 2023 年 5 月 24 日; 发布日期: 2023 年 5 月 31 日

摘 要

随着无人机电力巡检技术的发展, 无人机巡检逐渐替代了人工巡检, 成为输电线路监测和维护中不可或缺的工具。弧垂作为输电线路安全运行的关键指标, 准确的弧垂区间预测对于确保线路安全运行和及时维护具有重大意义。针对新型无人机巡检场景, 本研究提出了一种基于分位数极限梯度提升树 (QRGBDT) 的架空输电线路弧垂区间预测方法。通过将分位数回归模型与极限梯度提升树模型相结合, 构建了一个既能有效处理异常值、捕捉数据分布多样性, 又具有高度拟合能力和优异泛化性能的弧垂区间预测模型。通过实际应用案例以及和 QRCNN 和 QRRNN 模型的对比, 证实了本方法的有效性和优越性。

关键词

无人机巡检, 架空输电线路, 弧垂区间预测, 分位数极限梯度提升树

Power Line Sag Interval Prediction Based on QRGBDT for UAV Inspection Scenario

Yunfa Wu, Wenshuai Lin, Heng Duan, Jiahao Xie

School of Automation, Guangdong University of Technology, Guangzhou Guangdong

Received: Apr. 17th, 2023; accepted: May. 24th, 2023; published: May 31st, 2023

Abstract

With the development of drone power inspection technology, drone inspection has gradually replaced manual inspection, becoming an indispensable tool in the monitoring and maintenance of transmission lines. As a key indicator of the safe operation of transmission lines, accurate sag interval prediction is of great significance for ensuring the safe operation of lines and timely maintenance. In view of the new drone inspection scenario, this study proposes an overhead transmission line sag interval prediction method based on Quantile Regression Extreme Gradient Boosting Decision Tree (QRGBDT). By combining the quantile regression model with the extreme gradient boosting decision tree model, a sag interval prediction model is constructed that can effectively handle outliers, capture the diversity of data distribution, and has high fitting ability and excellent generalization performance. Through actual application cases and comparison with QRCNN and QRRNN models, the effectiveness and superiority of this method are confirmed.

Keywords

Drone Inspection, Overhead Transmission Lines, Sag Interval Prediction, QRGBDT

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着电力系统的发展和机器人技术的进步，机巡智能作业场景在输电线路的监测和维护中扮演越来越重要的角色 [1-3]。无人机激光雷达巡检具有高效、安全且成本较低的优势。然而，输电线路的分布具有较大的地理范围、复杂的地形和多样化的气候特点，在实际巡检过程中，可能会出现数据缺失、异常值和噪声等问题，影响巡检结果的准确性，这为电力无人机巡检数据的应用带来了巨大挑战 [4, 5]。

弧垂是衡量输电线路安全运行的关键指标之一，对于输电线路的运维和动态增容具有重要意义 [6-8]。通过对弧垂的变化，可以判断输电线路是否存在过热、过载等风险，及时采取相应措施防止事故的发生。准确的弧垂预测可以为运维人员提供更加精确的信息，以指导线路的维护工作。这有助于制定更为合理的运维计划，提高运维效率，降低运维成本。

实际应用中的数据往往受到诸多因素的影响，具有一定程度的不确定性 [9,10]。区间预测不仅给出预测值的点估计，而且提供了预测值的置信区间，使得预测结果更具有参考价值 [11]。准确的弧垂区间预测能够为线路的动态增容提供依据，可以准确判断线路的实际承载能力，从而为电力系统调度提供参考信息，实现电力资源的优化分配和更高效的利用。弧垂区间预测可以帮助实现预测性维护，可以在问题出现之前发现故障迹象，提前采取措施进行维修，从而提高输电线路的可靠性和稳定性。

区间预测已经成为众多领域的研究热点，涉及气象、经济、能源等多个领域 [12,13]。在弧垂预测方面，现有研究主要集中在传统统计方法、支持向量回归、神经网络等技术，这些方法在一定程度上取得了较好的预测效果 [14,15]。然而，由于输电线路环境的复杂性和数据的不确定性，现有方法在预测精度和鲁棒性方面仍有改进空间。

针对上述挑战，本文提出了一种基于分位数梯度提升树（QRGBDT）的架空输电线路弧垂区间预测方法，旨在为机巡智能作业场景提供精确、实时的弧垂预测。分位数回归具有对异常值具有较强鲁棒性和能够捕捉数据分布的多样性的特点，而梯度提升树（GBDT）则具有高度拟合能力和优异的泛化性能 [16]。该方法结合了分位数回归模型和极限梯度提升树模型的优点，构建了一个精确、鲁棒的弧垂区间预测模型。本文将对该方法进行详细介绍，并通过实际应用案例以及与其他模型的对比，验证其有效性和优越性。

2. 面向无人机巡检场景的弧垂数据集构建

本节将详细介绍面向无人机巡检场景的弧垂数据集构建过程，包括数据采集、基于知识的数据预处理以及数据集的构建。为了使模型能够更好地预测输电线路的弧垂，我们需要构建一个包含丰富特征的数据集。

2.1. 数据采集

数据采集是构建数据集的第一步。我们首先从无人机巡检过程中收集各类数据，这些数据包括输电线路的地理信息、气象条件、载荷等。这些数据是预测弧垂区间的基础。在本研究中，我们面向机巡作业特点及作业过程中可获得的相关数据情况，基于现有弧垂机理知识从机巡激光点云数据、作业天气数据、台账数据、导线信息数据等多源异构数据中提取与弧垂相关的有用信息，以帮助弧垂估计。

对于激光点云数据，提取每条线路的弧垂值作为预测模型的目标变量，提取导线悬挂点坐标计算线路档距、高差作为输入变量。对于作业天气数据，提取对弧垂变化有影响的天气因素，如环境温度、风速、风向。对于台账数据，提取对弧垂变化有影响的因素，如电压等级、导线型号、档类型、地形和服役时长等。对于导线参数，根据已有资料匹配导线型号对应的弹性系数、破断力、每千米电阻、导线直径、线膨胀系数、单位长度重量、总截面积、钢芯直径等。诸如太阳辐射

强度、风向、导线温度、线路电流等重要影响因素，受限于数据采集技术、部门间数据壁垒等原因无法获取。

2.2. 基于知识的数据预处理

在收集完原始数据后，我们需要对其进行清洗，以消除噪声、缺失值和异常值。清洗数据的目的是提高数据质量，从而使模型能够学习到更准确的关系。我们通过对弧垂悬链线机理模型知识、线路的基本状态方程知识、导线的热平衡方程知识的分析得到弧垂影响因素，从多源异构的数据中提取相关变量以获取有用信息，以期降低弧垂估计的不确定性。我们需要进行数据预处理，数据清洗的方法包括填充缺失值、剔除异常值等，还要进行特征提取和特征选择等步骤。

数据清洗：首先，我们需要对原始数据进行清洗，基于专家知识剔除不完整、错误或异常的数据记录，确保数据的清洗的合理性质量和可靠性。

特征提取：从无人机巡检数据中，结合弧垂悬链线机理模型知识、线路的基本状态方程知识、导线的热平衡方程知识的分析，我们可以提取一些有关弧垂的重要特征。例如，从激光点云数据中，我们可以提取导线档距、高差、弧垂、悬挂点位置、地形等。

特征重要性分析：重要性分析可以帮助我们了解不同特征对模型的预测结果的贡献程度。这对于模型的解释性和可解释性非常重要，因为它可以帮助我们理解模型是如何进行预测的。此外，特征重要性分析也可以帮助我们检查数据中的异常值、缺失值和噪声，以及对模型的鲁棒性和稳定性进行评估。

特征选择：特征重要性分析还可以帮助我们进行特征选择，即从模型中选择最重要的特征来提高模型的性能和效率。为了提高模型性能，我们需要从所有提取的特征中筛选出与弧垂预测最相关的特征。通过相关性分析、主成分分析（PCA）等方法实现。

经过数据预处理及特征选择后，我们基于搭载了激光雷达的无人机巡检作业过程，得到了以下特征如表 1。

2.3. 特征工程处理

在数据集构建之前，我们需要对原始数据进行特征工程处理。特征工程处理的目标是提取有意义的特征，以便更好地表示数据、降低数据的复杂性，从而提高模型的性能。针对无人机巡检作业场景下提取的数据，我们可以进行以下特征工程处理：

特征转换：对于可能存在非线性关系的特征，例如弹性系数（ELC）和破断力（BRF），可以尝试对数变换或平方根变换来观察它们与目标变量之间的关系是否得到改善。

特征编码：对于类别型特征，如电压等级（VTG）、档类型（SPT）和导线型号（CDT），需要将其转换为数值型特征。可以使用独热编码（One-Hot Encoding）或标签编码（Label Encoding）等方法。

特征缩放：对于数值型特征，如电线直径（DW）和总截面积（TCSA），可以进行归一化处理，以消除量纲和尺度的影响。可以选择最小-最大归一化（Min-Max Normalization）或标准化（Z-Score Normalization）。

Table 1. Sag prediction features and their abbreviations based on UAV inspection data**表 1.** 基于无人机巡检数据的弧垂预测特征及其缩写

数据类型	特征	英文全称	缩略词	类型
线路信息	电压等级	Voltage level	VTG	类别型
	档类型	Span type	SPT	类别型
	导线型号	Conductor type	CDT	类别型
	地形	Terrain	TR	类别型
	服役时长	Service time	ST	数值型
激光点云数据	档距	Span length	SPL	数值型
	高差	Height difference	HD	数值型
	最大弧垂	Maximum sag	MS	数值型
作业天气	环境温度	Ambient temperature	AT	数值型
	风速	Wind speed	WS	数值型
导线参数	弹性系数	Elastic coefficient	ELC	数值型
	破断力	Breaking force	BRF	数值型
	每千米电阻	Resistance per kilometer	RPK	数值型
	电线直径	Diameter of wire	DW	数值型
	线膨胀系数	Linear expansion coefficient	LEC	数值型
	单位长度重量	Weight per unit length	WPL	数值型
	总截面积	Total cross sectional area	TCSA	数值型
	钢芯直径	Steel core diameter	SCD	数值型

针对性地对这些特征进行特征工程处理后，我们所得到的数据将更具代表性，有利于构建更好的预测模型。

2.4. 数据集构建

经过数据预处理和特征工程后，我们得到一组包含丰富特征的数据记录。在构建数据集时，需要确保训练集和测试集的数据分布具有一定的代表性，以避免因数据分布不均导致的模型性能下降。此外，还需要对数据进行标准化处理，以消除不同特征之间的量纲影响，提高模型训练速度和性能。

为了便于模型训练和评估，我们需要将这些数据记录分为训练集和测试集。我们按照 80% 训练集、10% 测试集 10% 验证集的比例使用交叉验证的方法进行数据集划分。这样做的目的是为了在不同阶段评估模型的性能。验证集用于调整模型参数和超参数；测试集用于评估模型在未知数据上的泛化性能。

通过上述过程，我们构建了一个面向无人机巡检场景的弧垂数据集，包含了丰富的输电线路参数、环境因素和无人机巡检数据等特征。这为基于分位数梯度提升树的弧垂区间预测模型的训练和评估提供了有力支持。在后续实验和仿真验证中，我们将使用这个数据集来验证模型的性能。

3. 基于分位数极限梯度提升树的弧垂区间预测模型构建

在本节中，我们将详细介绍基于分位数极限梯度提升树（QRGBDT）的弧垂区间预测模型构建过程。首先，我们将简要回顾分位数回归模型和极限梯度提升树模型。接着，我们将介绍如何将这两种模型结合以构建弧垂区间预测模型。

3.1. 分位数回归模型

分位数回归（Quantile Regression, QR）是一种非参数统计方法，其目标是预测给定分位数下的条件分布。分位数回归模型可用于预测目标变量的不同分位数。给定观测值 X 和对应的目标变量 y ，分位数回归模型可以表示为：

$$y = X\beta(\tau) + \epsilon(\tau) \quad (1)$$

式中， $\beta(\tau)$ 表示第 τ 分位数的回归系数， $\epsilon(\tau)$ 表示第 τ 分位数的残差。分位数回归模型的目标是找到一组回归系数 $\beta(\tau)$ ，使得预测误差之和最小化：

$$\min_{\beta(\tau)} \sum_{i=1}^N \rho(\tau)(y_i - X_i\beta(\tau)) \quad (2)$$

式中， $\rho(\tau)(u) = u(\tau - I(u < 0))$ ， $I(\cdot)$ 是指示函数。

在分位数回归中，首先需要理解分位数损失函数。分位数损失函数（Quantile Loss Function）定义为：

$$L_\tau(u) = \rho_\tau(u) = u(\tau - I(u < 0)) \quad (3)$$

其中， u 是残差 ($y_i - X_i\beta(\tau)$)， τ 是我们关注的分位数（位于 $(0, 1)$ 区间内）， $I(\cdot)$ 是指示函数。这个损失函数在优化过程中对不同的残差赋予不同的权重。当残差为正时，权重为 τ ；当残差为负时，权重为 $(1 - \tau)$ 。

为了更好地理解这个损失函数的特性，我们可以观察其梯度。当残差 u 为正时，损失函数的梯度为 τ ；当残差 u 为负时，梯度为 $-\tau + 1$ 。这意味着较大的正残差受到较高的惩罚，而较大的负残差受到较低的惩罚。通过这种方式，分位数回归能够在不同分位数下捕捉目标变量的不同特征。

为了求解分位数回归模型，我们需要最小化如下目标函数：

$$\min_{\beta(\tau)} \sum_{i=1}^N L_\tau(y_i - X_i\beta(\tau)) = \sum_{i=1}^N \rho_\tau(y_i - X_i\beta(\tau)) \quad (4)$$

求解这个优化问题可以采用多种算法，如线性规划、内点法、梯度下降等。这些算法在计算过程中会迭代更新回归系数 $\beta(\tau)$ ，以最小化损失函数。

分位数回归模型的另一个有趣特性是它能够提供预测值的置信区间。通过计算目标变量在不

同分位数下的预测值，我们可以得到预测值的上下界。例如，对于一个 90% 的置信区间，我们可以计算第 5% 分位数和第 95% 分位数下的预测值，从而得到预测区间。这种置信区间估计在风险评估和决策支持中具有重要价值。

3.2. 梯度提升树模型

梯度提升树（Gradient Boosting Decision Tree, GBDT）是一种基于梯度提升方法的决策树集成算法，具有很强的拟合能力和泛化性能。GBDT 通过逐步训练一系列决策树，使得每棵树都针对前一棵树的残差进行拟合。设 $F_m(X)$ 表示第 m 棵树的预测结果，则 GBDT 模型可以表示为：

$$F(X) = F_0(X) + \sum_{m=1}^M \alpha_m F_m(X) \quad (5)$$

式中， $F_0(X)$ 是初始预测值， M 是树的数量， α_m 是第 m 棵树的权重。

GBDT 的关键在于每一步都使用损失函数的负梯度方向来构建新的决策树。在第 m 步，我们首先计算每个样本的负梯度值：

$$r_{im} = -\frac{\partial L(y_i, F_{m-1}(X_i))}{\partial F_{m-1}(X_i)} \quad (6)$$

其中， $L(y_i, F_{m-1}(X_i))$ 表示第 i 个样本在第 $m-1$ 步的损失值， r_{im} 是对应的负梯度。然后，我们使用这些负梯度值作为目标变量来拟合新的决策树 $F_m(X)$ 。

接下来，我们需要确定新的决策树 $F_m(X)$ 的权重 α_m 。这可以通过最小化以下损失函数来实现：

$$\alpha_m = \arg \min_{\alpha} \sum_{i=1}^N L(y_i, F_{m-1}(X_i) + \alpha F_m(X_i)) \quad (7)$$

通常情况下，我们可以使用线搜索（Line Search）来求解权重 α_m 。

最后，我们将新的决策树 $F_m(X)$ 乘以权重 α_m 并添加到当前模型中：

$$F_m(X) = F_{m-1}(X) + \alpha_m F_m(X) \quad (8)$$

GBDT 的训练过程持续进行 M 轮，直到满足停止条件（如最大迭代次数或损失值收敛）。最终的 GBDT 模型将由所有决策树的加权组合构成，从而在预测任务中表现出较强的拟合能力和泛化性能。

3.3. 分位数梯度提升树模型构建

结合分位数回归和梯度提升树模型，我们提出了分位数梯度提升树（QRGBDT）模型。该模型结合了分位数回归和极限梯度提升树的优点。首先，对于每个分位数 τ ，我们使用 GBDT 模型

来拟合分位数回归:

$$F(\tau)(X) = F_0(\tau)(X) + \sum_{m=1}^M \alpha_m(\tau) F_m(\tau)(X) \quad (9)$$

在训练过程中, 我们使用分位数回归损失来计算残差, 并调整模型参数以最小化损失函数:

$$\min_{\alpha_m(\tau), F_m(\tau)} \sum_{i=1}^N \rho(\tau)(y_i - F(\tau)(X_i)) \quad (10)$$

在分位数梯度提升树 (QRGBDT) 模型中, 我们在每个梯度提升步骤中针对分位数回归损失来计算负梯度。首先, 我们需要计算每个样本的负梯度值:

$$r_{im}(\tau) = -\frac{\partial \rho(\tau)(y_i - F_{m-1}(\tau)(X_i))}{\partial F_{m-1}(\tau)(X_i)} \quad (11)$$

其中, $\rho(\tau)(y_i - F_{m-1}(\tau)(X_i))$ 表示第 i 个样本在第 $m-1$ 步的分位数回归损失值, $r_{im}(\tau)$ 是对应的负梯度。

然后, 我们使用这些负梯度值作为目标变量来拟合新的决策树 $F_m(\tau)(X)$ 。在这个过程中, 我们可以使用分位数回归损失来优化每棵决策树的结构, 以提高模型在预测分位数时的准确性。

接下来, 我们需要确定新的决策树 $F_m(\tau)(X)$ 的权重 $\alpha_m(\tau)$ 。这可以通过最小化以下损失函数来实现:

$$\alpha_m(\tau) = \arg \min_{\alpha} \sum_{i=1}^N \rho(\tau)(y_i, F_{m-1}(\tau)(X_i) + \alpha F_m(\tau)(X_i)) \quad (12)$$

通常情况下, 我们可以使用线搜索 (Line Search) 来求解权重 $\alpha_m(\tau)$ 。

最后, 我们将新的决策树 $F_m(\tau)(X)$ 乘以权重 $\alpha_m(\tau)$ 并添加到当前模型中:

$$F_m(\tau)(X) = F_{m-1}(\tau)(X) + \alpha_m(\tau) F_m(\tau)(X) \quad (13)$$

QRGBDT 的训练过程持续进行 M 轮, 直到满足停止条件 (如最大迭代次数或损失值收敛)。最终的 QRGBDT 模型将由所有决策树的加权组合构成, 从而在预测任务中表现出较强的拟合能力和泛化性能。

3.4. 基于分位数梯度提升树的弧垂区间预测模型

为了构建弧垂区间预测模型, 我们首先针对给定的训练数据, 为每个分位数 τ 训练一个 QRGBDT 模型。然后, 我们将这些模型组合起来, 以预测给定输入特征的弧垂值在不同分位数上的预测区间。

给定一个输入特征向量 X_i ，我们可以通过如下方式计算预测区间：

$$\hat{y}_{i,\tau_1} = F(\tau_1)(X_i), \quad \hat{y}_{i,\tau_2} = F(\tau_2)(X_i) \quad (14)$$

式中， τ_1 和 τ_2 分别表示下分位数和上分位数。在预测阶段，我们可以利用这些模型为新的输电线路样本预测弧垂值的区间，从而为实际应用提供有关弧垂可能变化范围的信息。

置信区间（Confidence Interval）是一种用于表示参数估计不确定性的统计方法。在统计推断中，置信区间表示在给定置信水平下，未知参数的可能取值范围。置信区间在很多领域的数据分析中都有广泛应用，包括回归分析、假设检验和预测分析等。

对于一个总体参数 θ ，置信区间表示为：

$$(\hat{\theta}_L, \hat{\theta}_U) \quad (15)$$

式中， $\hat{\theta}_L$ 和 $\hat{\theta}_U$ 分别是置信区间的下限和上限。置信水平通常表示为 $(1 - \alpha)$ ，其中 α 是置信区间之外的概率，如 95% 置信水平对应的 α 值为 0.05。

在分位数回归的应用场景中，我们可以通过计算不同分位数下的条件分布来估计预测值的置信区间。例如，设 τ_1 和 τ_2 分别表示下分位数和上分位数，那么我们可以分别计算在这两个分位数下的预测值：

$$\hat{y}_{\tau_1} = X\beta_{\tau_1}, \quad \hat{y}_{\tau_2} = X\beta_{\tau_2} \quad (16)$$

这样，预测值 \hat{y} 在置信区间 $(\hat{y}_{\tau_1}, \hat{y}_{\tau_2})$ 内的概率为 $(1 - \alpha)$ 。通过计算置信区间，我们可以更准确地评估预测值的不确定性，从而在决策和风险评估过程中做出更明智的选择。

4. 基于贝叶斯优化的分位数梯度提升树区间预测模型

贝叶斯优化（Bayesian Optimization）是一种全局优化方法，它利用概率模型（通常是高斯过程）来找到目标函数的最优解。贝叶斯优化特别适用于高度非线性、昂贵计算且不易求导的函数优化问题。在超参数调优中，贝叶斯优化可以有效地搜索超参数空间并找到最佳超参数组合。

高斯过程（Gaussian Process, GP）是一种随机过程，其中任意有限个随机变量的联合分布是多元高斯分布。高斯过程可以表示为：

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')) \quad (17)$$

其中， $m(x)$ 是均值函数， $k(x, x')$ 是协方差函数。在贝叶斯优化中，我们使用高斯过程来建立目标函数的概率模型。

在贝叶斯优化中，我们首先设定一个先验分布，然后根据观测到的数据计算后验分布。给定目标函数 $f(x)$ ，我们可以通过以下公式计算后验分布：

$$p(f(x)|D) = \frac{p(D|f(x))p(f(x))}{p(D)} \quad (18)$$

其中, $D = (x_i, y_i)_{i=1}^n$ 是观测到的数据, $p(f(x)|D)$ 是后验分布, $p(D|f(x))$ 是似然函数, $p(f(x))$ 是先验分布, $p(D)$ 是边缘似然函数。

采集函数是一个用于平衡探索与利用的启发式函数。常见的采集函数有 Expected Improvement (EI), Probability of Improvement (PI) 和 Upper Confidence Bound (UCB)。例如, EI 采集函数可以表示为:

$$EI(x) = \mathbb{E}[f(x) - f(x^+)|D] \quad (19)$$

其中, x^+ 是目前已知的最优解。我们通过最大化采集函数来选择下一个采样点。

在基于贝叶斯优化的分位数梯度提升树区间预测模型的参数调优过程中, 我们首先设定一个超参数空间。然后, 我们采用以下迭代过程进行优化:

1. 初始化: 选取一组初始超参数, 计算目标函数值, 更新高斯过程模型。
2. 选择下一个采样点: 通过最大化采集函数来选择下一个超参数组合。
3. 评估目标函数: 使用新的超参数组合在训练集上训练分位数梯度提升树模型, 并计算验证集上的目标函数值 (如预测区间覆盖率、区间宽度等)。
4. 更新高斯过程模型: 将新的观测数据 (超参数组合和目标函数值) 纳入高斯过程模型, 并计算后验分布。判断收敛: 如果满足收敛条件 (如迭代次数达到预设值或目标函数的改善程度低于阈值), 则停止迭代; 否则, 返回第二步继续搜索。

通过以上迭代过程, 我们可以找到使目标函数最优化的超参数组合。这些超参数包括树的数量、树的深度、学习率等。找到最佳超参数组合后, 我们可以使用这些超参数在整个训练集上训练分位数梯度提升树模型, 从而得到一个强大的区间预测模型。

在基于贝叶斯优化的分位数梯度提升树区间预测模型中, 贝叶斯优化方法能够有效地搜索超参数空间并找到最佳超参数组合, 从而提高模型的预测性能。相较于传统的网格搜索和随机搜索方法, 贝叶斯优化方法能够更快地找到最优解, 从而节省了计算时间和资源。

5. 实验与仿真验证

本节将通过一个实验与仿真验证示例来展示基于分位数梯度提升树的弧垂区间预测模型的性能。我们将首先介绍实验设置和数据集, 然后展示实验结果, 并与其他方法进行比较。

5.1. 实验设置与数据集

我们使用某省若干条的输电线路弧垂相关的数据, 其中包含了 2310 条线路档的弧垂及其相关数据。每条数据包括表 1 中的特征。但是部分数据中存在数据缺失、噪声、误差大等情况, 如部分钢芯直径数据缺失、每千米电阻实际值与设计值误差大、环境温度和风速测量误差大、激光点云数据存在噪声, 因此我们的模型必须有较好的鲁棒性才能有效地处理以上问题。

我们将数据集按照 8:1:1 的比例划分为训练集、验证集和测试集。对于数值特征, 我们使用标准化方法进行预处理; 对于类别特征, 我们使用独热编码进行处理。

我们使用贝叶斯优化方法对分位数梯度提升树模型的超参数进行调优。在实验中，我们设定分位数梯度提升树模型的树的数量为 [10, 200]，树的最大深度为 [2, 10]，学习率为 [0.01, 0.3]。同时，我们选择预测区间覆盖率作为目标函数。

在完成数据集构建后，我们可以利用训练集来训练基于分位数梯度提升树的弧垂区间预测模型。在模型训练过程中，我们可以使用验证集来调整模型参数和超参数，以获得最佳性能。最后，我们可以使用测试集来评估模型在未知数据上的性能。

5.2. 特征重要性和相关性分析

弧垂数据集的特征重要性分析结果如图 1。

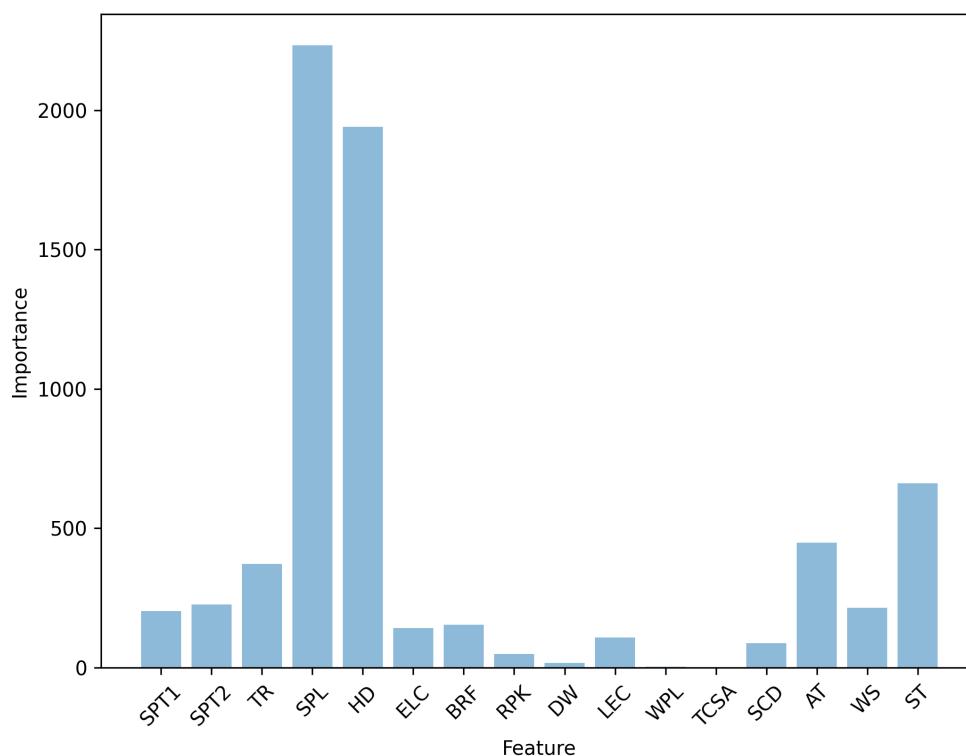


Figure 1. Feature importance analysis of sag dataset

图 1. 弧垂数据集的特征重要性分析

根据特征重要性结果，可以发现在这个模型中，档距（SPL）和高差（HD）是最重要的特征。而服役时间（ST）、环境温度（AT）、地形（TR）以及档塔的类型（SPT1, SPT2）也对预测结果有较大的影响。其余的特征对预测结果的影响相对较小。这些结果有助于我们进一步了解预测模型的表现，优化特征选择并改进模型性能。

特征的相关性分析结果如图 2 所示。

根据相关性矩阵结果，可以发现：

特征 0（SPT1）和特征 1（SPT2）之间的相关性很低（ -0.079 ），说明它们彼此独立；特征 3（SPL）和特征 4（HD）之间的相关性比较高（ 0.241 ），说明它们之间存在一定程度的相关性；特

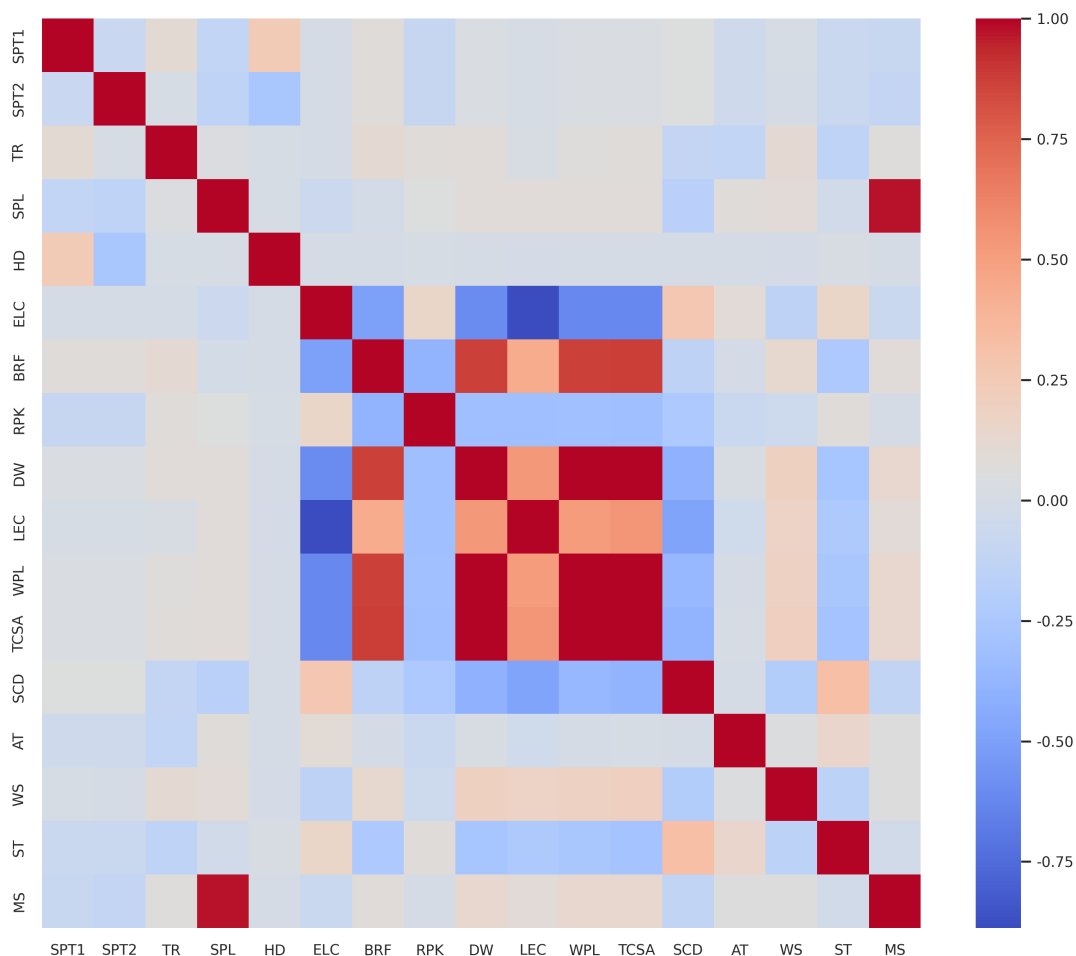


Figure 2. Correlation analysis of sag dataset

图 2. 弧垂数据集的相关性分析

特征 2 (TR) 与其他特征的相关性都比较低 (都小于 0.2), 说明它们彼此独立; 特征 5-12 之间的相关性较高, 其中特征 5 (ELC) 和特征 6 (BRF) 之间的相关性最高 (-0.503877), 说明它们之间存在负相关关系; 特征 15 (ST) 和特征 3 (SPL) 之间的相关性比较低 (-0.031748), 说明它们彼此独立; 特征 3 (SPL) 和特征 16 之间的相关性比较高 (0.974710), 说明它们之间存在一定程度的相关性。

根据上面的分析, 特征之间存在一定的相关性, 但是冗余度不高, 每个特征都有贡献, 因此我们选择将所有特征作为输入进行弧垂区间预测。

5.3. 区间预测性能指标

在评估区间预测模型的性能时, 我们选取了以下指标:

区间覆盖率 (Coverage Probability, CP): 区间覆盖率表示实际观测值落在预测区间内的比例。理想情况下, 区间覆盖率应接近置信水平。例如, 在 95% 置信水平下, 我们期望区间覆盖率

约为 95%。区间覆盖率计算公式如下：

$$CP = \frac{1}{N} \sum_{i=1}^N I(y_i \in [\hat{y}_i, \tau_1, \hat{y}_i, \tau_2]) \quad (20)$$

式中， N 是样本数量， $I(\cdot)$ 是指示函数， y_i 是实际观测值， \hat{y}_i, τ_1 和 \hat{y}_i, τ_2 分别表示预测区间的下界和上界。

平均区间宽度 (Mean Interval Width, MIW)：平均区间宽度表示预测区间的平均宽度，可以反映模型的不确定性。在保证区间覆盖率的前提下，我们希望平均区间宽度尽可能小。平均区间宽度计算公式如下：

$$MIW = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i, \tau_2 - \hat{y}_i, \tau_1) \quad (21)$$

通过以上性能指标，我们可以全面评估基于分位数梯度提升树的弧垂区间预测模型的性能，并与其他方法进行比较。

5.4. 弧垂区间预测结果

首先分析弧垂区间预测结果。在测试集上，我们使用训练好的分位数梯度提升树模型进行弧垂区间预测。我们选取 95% 的置信水平，得到的预测区间覆盖率为 94.7%，接近期望的 95%。同时，我们观察到预测区间的宽度较小，说明模型在保证预测精度的同时，能够给出较为紧凑的预测区间。

我们将基于分位数梯度提升树的弧垂区间预测模型与其他区间预测方法进行比较，选择分位数卷积神经网络 (QRCNN) 和分位数循环神经网络 (QRRNN) 作为对比方法进行比较。为了更可靠地评估模型性能，我们使用 10-折交叉验证方法。交叉验证可以减小模型性能评估的方差，提高评估结果的稳定性。如图 3 为模型在弧垂区间预测上的效果。

结果显示，分位数梯度提升树模型在预测区间覆盖率和区间宽度方面均优于其他方法。如表 2 为各模型的性能指标，包括均方误差、平均绝对误差、 R^2 得分、平均区间宽度和区间覆盖率结果。

Table 2. Performance comparison of sag interval prediction among QRGBDT, QRCNN, and QRRNN

表 2. QRGBDT、QRCNN、QRRNN 在弧垂区间预测上的效果

模型名称	均方误差 (MSE)	平均绝对误差 (MAE)	R^2 得分	MIW	CP
QRRNN	1.0329	0.8250	-1.6439	0.0	0.0
QRCNN	0.7372	0.6968	-0.8870	0.5466	0.1797
QRGBDT	0.0361	0.1433	0.9076	0.2722	0.8485

根据表 2 中的指标，可以看出 QRGBDT 模型在弧垂区间预测方面的性能表现最优，其均方误差 (MSE) 为 0.0361，平均绝对误差 (MAE) 为 0.1433， R^2 得分为 0.9076，MIW 为 0.2722，CP 为 0.8485。这些指标都比另外两个模型 (QRRNN 和 QRCNN) 要好。

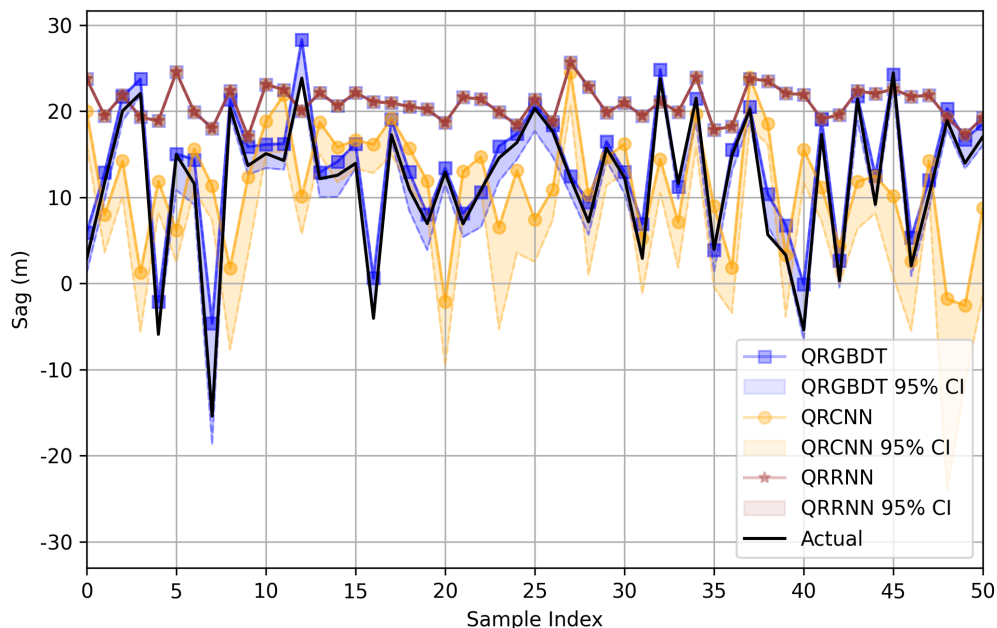


Figure 3. Sag interval prediction of the model

图 3. 模型的弧垂区间预测

QRRNN 模型的均方误差 (MSE) 为 1.0329, 平均绝对误差 (MAE) 为 0.8250, R^2 得分为 -1.6439 , MIW 为 0.0, CP 为 0.0。其中 R^2 得分为负数, 表明模型的表现不如直接用平均值进行预测, 预测结果偏差较大, 区间预测效果差。

QRCNN 模型的均方误差 (MSE) 为 0.7372, 平均绝对误差 (MAE) 为 0.6968, R^2 得分为 -0.8870 , MIW 为 0.5466, CP 为 0.1797。虽然性能较 QRRNN 模型有所提升, 但与 QRGBDT 模型相比, 仍有较大差距。

因此, QRGBDT 模型在所有指标上表现最优, 其均方误差和平均绝对误差都远低于其他两个模型, 且 R^2 得分远高于其他两个模型。同时, MIW 和 CP 也是最好的。因此, 我们可以认为 QRGBDT 模型在弧垂区间估计上表现最佳。

5.5. 贝叶斯优化结果

Table 3. Hyperparameter selection range and optimal hyperparameters results of Bayesian optimization

表 3. 贝叶斯优化的超参数选择范围及最优超参数结果

超参数	选取范围	最优超参数
学习率	0.01~0.2	0.1923
最大深度	3~8	6.7938
最小叶子样本数	10~50	20.8391
子采样率	0.5~0.9	0.5367
列采样率	0.5~0.9	0.8188

我们使用贝叶斯优化器优化 QRGBDT 模型的超参数，以提高 QRGBDT 的弧垂区间预测性能。我们首先定义了超参数搜索范围，包括学习率、最大深度、最小叶子样本数、子采样率和列采样率等。然后我们定义了需要优化的目标函数，即在给定超参数下训练模型并返回其均方误差。接着我们使用 Bayesian Optimization 来搜索最优的超参数组合，并使用该组合训练模型。最终，我们输出了模型的均方误差、平均绝对误差、 R^2 得分、MIW 得分和 CP 得分等评估指标，以评估模型性能。Bayesian 优化器找到了一组最优超参数，表 3 为我们设置的超参数选取范围和最优超参数结果。

接下来需要使用这组最优超参数重新训练 QRGBDT 模型，并评估其性能。如图 4 为贝叶斯优化后的模型弧垂区间预测效果。

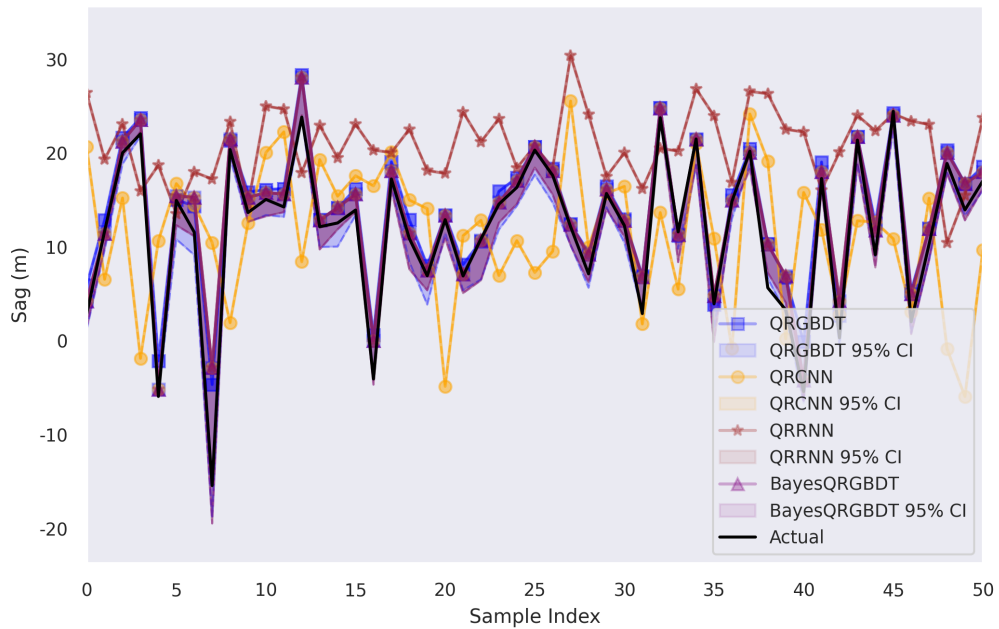


Figure 4. Sag interval prediction performance of models after Bayesian optimization

图 4. 贝叶斯优化后的模型弧垂区间预测效果

表 4 为 BayesQRGBDT 和 QRGBDT 模型性能对比。

Table 4. Performance comparison between BayesQRGBDT and QRGBDT models

表 4. BayesQRGBDT 和 QRGBDT 模型性能对比

	BayesQRGBDT	QRGBDT
MSE	0.0319	0.0361
MAE	0.1261	0.1433
R^2 Score	0.9183	0.9076
MIW	0.2287	0.2722
CP	0.7706	0.8485

可以看到，使用 Bayesian 优化的 QRGBDT 相较于普通的 QRGBDT 表现更好。具体来说，

它的均方误差 (MSE)、平均绝对误差 (MAE)、MIW 和 CP 都表现更好, 而 R^2 得分也更高。这表明使用 Bayesian 优化的 QRGBDT 在预测上更加准确和可靠, 优化过程可以帮助我们找到最优的超参数组合来提高模型的性能。

6. 结论

本文针对面向无人机巡检作业场景, 使用 QRGBDT 模型预测架空输电线路弧垂区间。首先, 我们使用传统的 QRGBDT 模型进行训练和预测, 得到的预测结果表明该模型可以预测架空输电线路弧垂区间, 并且具有较高的预测准确度。但是, 这种传统的 QRGBDT 模型无法自适应地根据数据分布调整分位数的值, 而且容易受到异常值的干扰。

因此, 我们引入了 BayesQRGBDT 模型, 使用贝叶斯优化方法优化超参数, 可以自适应地根据数据分布调整分位数的值, 从而提高了预测的鲁棒性和准确性。实验结果表明, BayesQRGBDT 模型相对于传统的 QRGBDT 模型, 可以更好地预测架空输电线路弧垂区间。

综上所述, 本文提出的基于 QRGBDT 的架空输电线路弧垂区间预测方法, 可以在无人机巡检作业场景中应用, 并取得较好的预测效果, 可以为电力运营管理提供较为精准的弧垂区间预测结果, 进一步提高输电线路的安全性和可靠性。

基金项目

广东电网科技项目 (GDKJXM20184755)。

参考文献

- [1] da Silva, M.F., Honório, L.M., Marcato, A.L.M., Vidal, V.F. and Santos, M.F. (2020) Unmanned Aerial Vehicle for Transmission Line Inspection Using an Extended Kalman Filter with Colored Electromagnetic Interference. *ISA Transactions*, **100**, 322-333. <https://doi.org/10.1016/j.isatra.2019.11.007>
- [2] 彭向阳, 易琳, 钱金菊, 王柯, 郑晓光, 韩正伟, 陈国强. 大型无人直升机电力线路巡检系统实用化 [J]. *高电压技术*, 2020, 46(2): 384-396.
- [3] Zhang, Y., Yuan, X.X., Li, W.Z. and Chen, S.Y. (2017) Automatic Power Line Inspection Using UAV Images. *Remote Sensing*, **9**, Article 824. <https://doi.org/10.3390/rs9080824>
- [4] 常龙, 游华武. 无人机激光雷达技术在高压输电线路三维设计中的应用 [J]. *电声技术*, 2021, 45(10): 80-82.
- [5] 隋宇, 宁平凡, 牛萍娟, 王辰羽, 赵地, 张伟龙, 韩抒真, 梁立君, 薛高建, 崔颜军. 面向架空输电线路的挂载无人机电力巡检技术研究综述 [J]. *电网技术*, 2021, 45(9): 3636-3648.
- [6] 全国架空线路标准化技术委员会线路运行分技术委员会. DL/T 741-2019. 架空输电线路运行规程 [S]. 北京: 中国电力出版社, 2019: 1-36.

-
- [7] IEEE (2013) IEEE Standard for Calculating the Current-Temperature Relationship of Bare Overhead Conductors. IEEE Std 738-2012, IEEE, New York, 1-58.
- [8] 孟遂民, 孔伟. 架空输电线路设计 [M]. 第 2 版. 北京: 中国电力出版社, 2015.
- [9] Xu, Y.J., Huang, C., Chen, X., Mili, L., Tong, C.H., Korkali, M. and Min, L. (2019) Response-Surface-Based Bayesian Inference for Power System Dynamic Parameter Estimation. *IEEE Transactions on Smart Grid*, **10**, 5899-5909. <https://doi.org/10.1109/TSG.2019.2892464>
- [10] Ye, G., Xiang, Y., Nijhuis, M., Cuk, V. and Cobben, J.F.G. (2017) Bayesian-Inference-Based Voltage Dip State Estimation. *IEEE Transactions on Instrumentation and Measurement*, **66**, 2977-2987. <https://doi.org/10.1109/TIM.2017.2734138>
- [11] Wang, Y., Zhou, Z., Botterud, A. and Zhang, K.F. (2017) Optimal Wind Power Uncertainty Intervals for Electricity Market Operation. *IEEE Transactions on Sustainable Energy*, **9**, 199-210. <https://doi.org/10.1109/TSTE.2017.2723907>
- [12] Yue, C.-D., Chiu, Y.-S., Tu, C.-C. and Lin, T.-H. (2020) Evaluation of an Offshore Wind Farm by Using Data from the Weather Station, Floating LiDAR, MAST, and MERRA. *Energies*, **13**, Article 185. <https://doi.org/10.3390/en13010185>
- [13] Wang, S.J., Zeng, J.Y. and Liu, X.P. (2019) Examining the Multiple Impacts of Technological Progress on CO₂ Emissions in China: A Panel Quantile Regression Approach. *Renewable and Sustainable Energy Reviews*, **103**, 140-150. <https://doi.org/10.1016/j.rser.2018.12.046>
- [14] Zheng, H.T., Yuan, J.B. and Chen, L. (2017) Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies*, **10**, Article 1168. <https://doi.org/10.3390/en10081168>
- [15] Jia, X.W., Willard, J., Karpatne, A., Read, J.S., Zwart, J.A., Steinbach, M. and Kumar, V. (2021) Physics-Guided Machine Learning for Scientific Discovery: An Application in Simulating Lake Temperature Profiles. *ACM/IMS Transactions on Data Science*, **2**, Article No. 20. <https://doi.org/10.1145/3447814>
- [16] Chen, T.Q., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K.L., *et al.* (2015) XGBoost: Extreme Gradient Boosting. <https://cran.r-project.org/web/packages/xgboost/vignettes/xgboost.pdf>