基于大数据分析的铁路供电接触网故障预测模型研究

郭武杰

国能黄大铁路有限责任公司 山东 东营 257000

摘要: 随着货运铁路重载化发展,铁路供电接触网作为保障列车安全运行的关键基础设施,其稳定性和可靠性至关重要。然而,接触网结构复杂、工作环境恶劣,故障频发,给铁路运营带来了严峻挑战。本文旨在利用大数据分析技术,构建铁路供电接触网故障预测模型,实现对接触网故障的提前预警和精准预测,为铁路运维提供科学依据。

关键词: 大数据;铁路供电接触网;故障预测模型

引言

铁路供电接触网是电气化铁路的重要组成部分,负责为列车提供电能。由于其长期处于户外恶劣环境中,且结构复杂、零部件众多,故障时有发生。传统的接触网检修方式主要依赖于定期检测和故障后维修,存在检测不及时、维修成本高、影响运营效率等问题。因此,利用大数据分析技术构建接触网故障预测模型,实现故障的早期预警和精准预测,对于提高铁路供电系统的可靠性和安全性具有重要意义。

1 接触网故障特征分析

接触网故障类型多样,包括短路故障、断线故障、绝缘子故障等。这些故障的发生往往与接触网的运行环境、设备状态、维护质量等多种因素有关。通过对大量接触网故障数据的统计分析,可以发现不同故障类型具有不同的特征模式。

1.1 短路故障特征

短路故障是接触网较为常见的故障之一,其主要原因包括接触线与承力索短路、绝缘子击穿等。在短路故障发生时,故障点的电流会急剧增大,通常会超过正常工作电流的数倍甚至数十倍。例如,在正常运行情况下,接触网的电流可能稳定在几百安培,而发生短路故障时,电流可能会瞬间飙升至几千安培甚至更高。同时,电压也会出现明显的波动,故障点附近的电压可能会瞬间降低,导致供电质量下降。此外,短路故障还可能引发电弧,电弧产生的高温会对接触网设备造成进一步的损坏,如烧蚀接触线、损坏绝缘子等。

1.2 断线故障特征

断线故障可能是由于接触线或承力索的机械磨损、疲劳断裂等原因引起的。当发生断线故障时,供电会立即中断,受电弓与接触网之间的接触状态会发生改变,导致受电弓磨损加剧。具体来说,受电弓在失去与接触线的正常接触后,会在断线处产生剧烈的跳动和摩擦,

从而加速受电弓滑板的磨损。同时,断线部分可能会在风力等外力作用下摆动,对周围的设备和线路造成威胁,如与相邻的线索发生碰撞,引发新的故障。

1.3 绝缘子故障特征

绝缘子故障主要表现为绝缘性能下降或击穿。绝缘子长期暴露在恶劣的环境中,会受到污秽、潮湿等因素的影响,导致其绝缘性能逐渐降低。当绝缘子的绝缘性能下降到一定程度时,就可能发生击穿故障,引发短路等严重后果。绝缘子故障的特征可能包括泄漏电流增大、表面闪络等现象。泄漏电流是指在正常工作电压下,绝缘子表面流过的微弱电流。当绝缘子性能下降时,泄漏电流会明显增大。表面闪络则是指在潮湿、污秽等条件下,绝缘子表面出现局部放电现象,严重时会导致绝缘子击穿。

2 大数据分析在接触网故障预测中的应用

2.1 数据采集与整合

接触网故障预测模型的构建首先需要大量的数据支持。这些数据包括接触网的运行状态数据(如电流、电压、温度等)、环境数据(如风速、湿度、温度等)、历史故障记录等。通过传感器网络、物联网技术等手段,可以实现对这些数据的实时采集和传输。

在接触网沿线安装各类传感器,如电流传感器、电压传感器、温度传感器等,用于实时监测接触网的运行状态。电流传感器采用霍尔效应原理,能够精确测量接触网中的电流大小;电压传感器则通过分压原理,将高电压转换为低电压信号进行测量。温度传感器采用热敏电阻或热电偶等元件,实时感知接触网设备的温度变化。同时,利用气象站等设备获取环境数据,气象站可以测量风速、风向、湿度、温度、气压等参数^[1]。这些数据通过无线通信网络传输到数据中心进行存储和管理。常见的无线通信网络包括GPRS、ZigBee、LoRa等,它们具有不同的特点和适用场景。

在数据采集过程中,需要确保数据的准确性和完整性,避免因传感器故障、通信中断等原因导致数据丢失或错误。为此,需要定期对传感器进行校准和维护,检查通信线路的连接情况。对采集到的数据进行整合和清洗是关键步骤。由于数据来源广泛、格式多样,需要进行数据格式转换、缺失值处理、异常值剔除等操作。例如,对于不同传感器采集到的数据,可能采用不同的数据格式,需要将其统一转换为便于处理的格式。对于缺失的数据,可以采用插值法、均值法等方法进行填充。插值法是根据已知数据点的趋势,估算出缺失数据点的值;均值法则是用该数据列的平均值来填充缺失值。对于异常值,可以通过统计分析、机器学习算法等方法进行识别和剔除。统计分析方法如 3σ 原则,即当数据点偏离均值超过3倍标准差时,认为该数据点为异常值。

2.2 特征提取与选择

在构建预测模型之前,需要从原始数据中提取出对预测目标有影响的特征变量。这些特征变量可能包括接触网的电气参数、机械参数、环境参数等。

对于电气参数,可以提取电流的有效值、峰值、谐波含量等特征。电流的有效值反映了电流的平均功率,峰值则体现了电流的最大瞬时值。谐波含量是指电流中除基波频率外的其他频率成分的含量,谐波的存在会对电力系统的运行产生不良影响。通过傅里叶变换等方法可以对电流信号进行分析,提取出这些特征。对于机械参数,可以提取接触线的张力、振动频率等特征。接触线的张力是影响接触网稳定性的重要因素,张力过小可能导致接触线松弛,张力过大则可能加速接触线的磨损。振动频率与接触网的运行状态和外部环境有关,异常的振动频率可能预示着接触网存在故障隐患。可以通过安装张力传感器和振动传感器来测量这些机械参数。对于环境参数,可以提取风速、湿度、温度等特征。这些环境参数会对接触网的运行状态产生影响,如风速过大会导致接触线晃动加剧,湿度过高会加速绝缘子的老化。

特征选择是为了从提取的特征中选择出对预测结果最有影响的特征,减少数据的维度,提高预测的准确性和效率。常用的特征选择方法包括过滤法、包裹法和嵌入法等。过滤法是根据特征的统计特性进行筛选,如相关系数、方差等。相关系数可以衡量特征与目标变量之间的线性关系,方差则可以反映特征的离散程度^[2]。通过设定一定的阈值,筛选出相关系数较高、方差较大的特征。包裹法是通过构建模型来评估特征的重要性,例如采用逐步回归的方法,每次加入或删除一个特征,观察模型性能的变化,从而确定最优的特征子集。嵌入法是

在模型训练过程中同时进行特征选择,如决策树算法在构建树的过程中,会根据特征的信息增益来选择分裂节点,从而实现特征的选择。

2.3 模型构建与训练

基于提取的特征变量,可以构建接触网故障预测模型。常用的预测模型包括支持向量机(SVM)、神经网络(NN)、决策树(DT)等。

支持向量机是一种基于统计学习理论的分类和回归方法,它通过寻找最优的超平面来划分不同类别的数据。在接触网故障预测中,可以将正常状态和故障状态作为不同的类别,利用支持向量机进行分类预测。支持向量的关键在于核函数的选择,常见的核函数有线性核、多项式核、径向基核等。不同的核函数适用于不同类型的数据分布,需要根据实际情况进行选择。

神经网络具有强大的非线性映射能力,能够自动学习数据中的复杂模式。在接触网故障预测中,可以采用多层感知器(MLP)、卷积神经网络(CNN)等神经网络模型。多层感知器由输入层、隐藏层和输出层组成,通过调整神经元之间的连接权重来实现对数据的分类和预测。在训练过程中,采用反向传播算法来调整权重,使得模型的输出与实际标签之间的误差最小。卷积神经网络则更适用于处理具有空间结构的数据,如接触网的图像数据。它通过卷积层、池化层和全连接层的组合,自动提取图像中的特征,并进行分类预测。

决策树是一种基于树结构进行决策的模型,它通过对数据进行划分,构建决策规则来进行预测。决策树具有可解释性强的优点,能够直观地展示决策过程。在构建决策树时,需要选择合适的划分属性,常用的划分标准有信息增益、增益率等。

在模型训练过程中,需要将数据集分为训练集和测试集。通常采用交叉验证的方法,将数据集划分为多个子集,轮流将其中一个子集作为测试集,其余子集作为训练集,多次训练和测试模型,以评估模型的稳定性和泛化能力。利用训练集对模型进行训练,调整模型的参数,使模型能够学习到数据中的规律^[3]。然后,利用测试集对训练好的模型进行评估,计算模型的准确率、召回率、F1值等指标。准确率是指模型预测正确的样本数占总样本数的比例;召回率是指模型正确预测出的正样本数占实际正样本数的比例;F1值则是准确率和召回率的调和平均数,综合考虑了模型的精确性和全面性。

2.4 故障预测与预警

训练好的模型可以用于对接触网故障进行预测。通过将实时监测到的数据输入到预测模型中,可以得到接

触网故障的可能性和发生时间等信息。当预测结果显示故障可能性较高时,可以提前采取措施进行预警和维修,避免故障发生和影响列车运行。

预警系统可以根据预测结果自动发出警报,通知运维人员及时进行检查和维修。警报方式可以包括短信、邮件、声光报警等。同时,预警系统还可以提供故障的可能位置和类型等信息,帮助运维人员快速定位和处理故障。例如,当预测到某段接触线可能发生断线故障时,预警系统会准确给出该段接触线的位置信息,并提示运维人员准备相应的维修工具和材料。

3 未来优化方向

未来,随着大数据技术和人工智能技术的不断发展,接触网故障预测模型将更加智能化和精准化。以下是一些未来的研究方向:

3.1 优化数据采集和整合流程

在数据采集方面,未来的研究将更加注重提高数据的准确性和实时性。当前的数据采集系统虽然已能基本满足需求,但在面对海量、高速生成的数据时,仍存在一定的延迟和误差。因此,探索更高效的数据采集技术,如利用物联网(IoT)设备实现数据的实时传输,以及采用更高精度的传感器提高数据质量,将成为重要研究方向。同时,数据整合和清洗也是数据处理过程中的关键环节。面对海量数据,如何快速、有效地进行整合和清洗,去除冗余和错误数据,提高数据质量,是亟待解决的问题。未来,可以借鉴分布式存储和计算技术,如Hadoop、Spark等,实现数据的高效处理。

3.2 提高特征提取和选择效率

特征提取和选择是构建预测模型的关键步骤。未来的研究将致力于开发更加先进的特征提取算法,以提高特征的代表性和可解释性。例如,可以引入深度学习技术,如卷积神经网络(CNN)或循环神经网络(RNN),自动学习数据中的高级特征,减少人工特征工程的工作量。此外,特征选择方法也需要进一步优化。当前的特征选择方法往往计算复杂度较高,难以处理大规模数据集。因此,开发更高效的特征选择算法,如基于贪心策略的特征选择方法或基于随机森林的特征重要性评估方法,将有助于提高模型的训练速度和预测性能。

3.3 增强模型的泛化能力

接触网故障预测模型需要适应不同铁路线路的特点和差异,因此增强模型的泛化能力至关重要。未来的研究可以考虑引入迁移学习方法,将在一个线路上训练好的模型迁移到另一个线路上进行微调,以适应新环境的数据分布。此外,还可以通过集成学习等方法,结合多个模型的预测结果,提高整体预测的准确性和稳定性。

3.4 融合多源数据

除了接触网的运行状态数据和环境数据外,还可以融合更多源的数据来提高预测的准确性和可靠性。例如,列车运行数据(如速度、加速度等)可能与接触网故障存在一定的关联性;设备维护记录也可以为预测模型提供有价值的先验信息。未来的研究将探索如何有效地融合这些多源数据,构建更加全面的预测模型。

3.5 实现智能运维决策

最终,接触网故障预测模型的目的是为铁路运维提供决策支持。因此,将预测模型与智能运维决策系统相结合,根据预测结果自动生成最优的运维方案,是未来的重要研究方向。这包括根据故障的类型和位置,合理安排维修人员和物资;以及根据预测结果制定长期的维护计划,优化资源分配和降低运维成本。通过实现智能运维决策,可以显著提高铁路运维的效率和效益,保障铁路运输的安全和稳定。

结语

本文通过对接触网故障特征的分析,明确了不同故障类型的特征模式。利用大数据分析技术,实现了对接触网运行状态数据、环境数据和历史故障记录的有效采集、整合和清洗。通过特征提取与选择,提高了模型的预测准确性和效率。总之,基于大数据分析的铁路供电接触网故障预测模型具有广阔的应用前景和发展潜力,需要不断地进行研究和改进,以更好地保障铁路供电系统的安全稳定运行。

参考文献

- [1]潘卫波.大数据技术在接触网供电故障分析中的应用[J].智能城市,2021,7(06):65-66.
- [2]李宏峰.基于大数据平台的接触网健康管理[J].电气化铁道,2020,31(S2):50-53.
- [3]李欣,白兴武.基于大数据分析的接触网道岔自动检修装置[J].自动化与仪器仪表,2020,(05):102-105.