

火电厂热工测量数据预处理方法研究

邹振宇 张海亮 杨思珮 孙艳莉

内蒙古上都发电有限责任公司 内蒙古 锡林郭勒盟 027200

摘要：火电厂热工测量数据对设备运行与决策至关重要，但具有实时性强、多源性、非线性和时变性等特点，且存在噪声、缺失值和异常值等问题。本文介绍了噪声处理的均值、中值和小波滤波方法，缺失值处理的插值、时间序列预测和机器学习补全策略，以及异常值处理的统计、基于距离和深度学习检测方法，旨在提高数据质量，助力火电厂智能化运行。

关键词：火电厂；热工测量数据；数据预处理；研究

引言：在火电厂复杂且高要求的生产运行体系里，热工测量数据意义重大，它是设备运行状态的直观反映，更是精准控制与优化决策的关键依据。然而，火电厂热工测量数据具有实时性强、多源性、非线性和时变性等独特特点，且实际采集时易受传感器老化、电磁干扰、设备故障等因素影响，出现噪声、缺失值和异常值等问题，严重影响数据可用性。因此，对热工测量数据进行有效的预处理至关重要。本文将围绕火电厂热工测量数据的噪声、缺失值和异常值处理方法展开深入探讨。

1 火电厂热工测量数据特点与预处理需求

在火电厂的复杂生产体系中，热工测量数据扮演着至关重要的角色，它不仅是反映设备运行状态的关键指标，更是实现精准控制与优化决策的重要依据。然而，火电厂热工测量数据具有一系列独特的特点，同时也面临着诸多预处理需求。（1）火电厂热工测量数据实时性极强。生产过程中，各类传感器以高频次持续采集数据，这些数据如同电厂运行的“脉搏”，实时反映着设备的运行状况。海量时间序列数据由此生成，为后续分析提供了丰富的信息源，但同时也对数据存储、处理和数据分析能力提出了极高要求。（2）数据多源性是另一显著特点。火电厂包含锅炉、汽轮机、发电机等多个复杂系统，每个系统都有众多监测参数，如温度、压力、流量等。这些参数来自不同的传感器和监测设备，数据来源广泛且相互关联，共同构成了火电厂热工测量数据的庞大网络。（3）热工测量数据还呈现出非线性和时变性特征。设备运行过程中，各参数之间的关系并非简单的线性关系，而是相互影响、相互制约，存在复杂的耦合关系。同时，随着设备运行时间的推移、工况的变化以及外部环境因素的影响，数据也会随之发生动态变化。（4）实际采集到的热工测量数据往往存在诸多问题。传感器老化、信号传输过程中的电磁干扰等因素，会导

致数据出现波动噪声，使数据偏离真实值；设备故障或通信中断则可能造成数据缺失，影响数据的完整性和连续性；操作失误、仪表故障等还可能产生异常值，干扰对设备运行状态的准确判断。这些问题严重影响了热工测量数据的可用性，若不进行妥善处理，将直接导致后续数据分析结果不准确，控制决策失误，进而影响火电厂的安全稳定运行和经济性。因此，数据预处理成为火电厂热工测量数据处理的关键环节。通过运用滤波算法去除噪声、采用插值或模型预测方法填补缺失值、利用统计方法或机器学习算法识别并修正异常值等预处理技术，能够有效提高热工测量数据的质量，为后续的数据分析、故障诊断、性能评估以及控制决策提供可靠的数据基础，助力火电厂实现智能化、高效化运行^[1]。

2 火电厂热工测量数据噪声处理方法

2.1 滤波算法原理与应用

均值滤波作为线性滤波领域中一种简单且行之有效的方法，其核心原理在于对数据窗口内的数据进行平均计算，以此实现数据的平滑处理，进而抑制随机噪声。具体而言，在给定的数据序列中，选取一个固定大小的窗口，将窗口内所有数据值相加后除以窗口内的数据个数，得到的结果即为该窗口中心位置数据的新值。通过不断移动窗口，对整个数据序列进行处理，可有效降低数据的随机波动。在火电厂热工测量数据处理场景中，对于温度、压力等随时间变化较为缓慢的参数，均值滤波能够发挥显著作用。以锅炉炉膛温度测量数据处理为例，炉膛温度在正常运行时虽有波动，但整体趋势相对平稳。此时设置合适的窗口大小，对连续采集的温度数据进行均值滤波处理，能使温度曲线更加平滑，消除小幅波动噪声的干扰，从而更准确地判断温度的变化趋势，为后续的锅炉运行控制提供可靠依据。

2.2 中值滤波技术分析

中值滤波属于非线性滤波方法，其独特之处在于它并非基于数据的线性组合来获取滤波结果。其具体操作流程为：首先确定一个数据窗口，将窗口内的所有数据按照数值大小进行排序，随后选取排序后位于中间位置的数据值作为该窗口中心位置的滤波输出值。通过不断滑动窗口，对整个数据序列进行处理。在火电厂热工测量领域，由于现场存在复杂的电磁环境，测量数据中常常会混入脉冲型噪声。中值滤波在处理这类噪声时展现出显著优势。以汽轮机油压测量数据为例，当受到电磁干扰时，数据中可能会出现瞬间尖峰噪声。这些尖峰噪声会严重偏离真实油压值，若不进行处理，极易引发油压异常报警误触发。而中值滤波能够快速识别并剔除这些脉冲噪声，精准保留真实的油压变化趋势，从而为汽轮机的安全稳定运行提供可靠的数据支持，保障火电厂的生产安全与效率^[2]。

2.3 小波滤波方法研究

小波滤波是建立在小波变换理论基础之上的一种先进滤波方法。小波变换具有多分辨率分析的特性，它能够将信号分解到不同的频率尺度上，将信号中的各种频率成分清晰地呈现出来。在信号分解过程中，信号被分解为一系列小波系数，这些系数分别对应着不同频率段的信号特征。在火电厂复杂多变的工业环境下，热工测量数据往往会受到多种频率成分噪声的干扰。小波滤波正是利用信号与噪声在不同频率段所表现出的特性差异来发挥作用。它通过对小波系数进行处理，有针对地去掉噪声频段所对应的小波系数，同时保留有用信号成分的小波系数。以蒸汽流量测量数据为例，该数据中既包含反映流量真实变化趋势的低频成分，又混杂着高频噪声。小波滤波能够精准地将高频噪声与低频趋势成分分离，在去除噪声的同时，完整保留流量变化的细节特征，从而确保蒸汽流量数据的准确性，为火电厂的流量监测与控制提供可靠依据。

3 火电厂热工测量数据缺失值处理方法

3.1 插值补全算法

线性插值法作为插值补全算法中的基础方法，在处理缺失值问题上具有简单易行的特点。其基本思路是依据缺失值前后相邻的两个已知数据点，假定这两个点之间的数据变化呈线性关系，进而估算出缺失位置的数据值。在火电厂热工测量数据处理中，线性插值法有着特定的应用场景。例如，当某段时间内部分温度数据出现缺失，且经过对历史数据和现场工况的分析，判断温度变化趋势近似线性时，就可以采用线性插值法进行补全。这种方法计算过程相对简便，不需要复杂的数学模

型和大量的计算资源，能够在短时间内完成缺失值的填补。然而，线性插值法也存在明显的局限性。当实际数据变化波动较大，并非呈现线性趋势时，该方法所补全的数据与真实值之间会存在较大误差，补全精度较低，可能无法准确反映数据的真实情况，进而影响后续对热工测量数据的分析和决策^[3]。

3.2 时间序列预测补全

基于时间序列分析的预测补全方法在处理数据缺失问题上具有独特优势，常见的模型包括自回归滑动平均模型（ARMA）、季节性分解法等。这些方法的核心在于深入剖析历史数据的内在特征，如数据随时间变化的趋势、周期性波动等，以此为基础构建精准的预测模型，进而对缺失值进行合理估算。在火电厂的实际运行中，许多热工参数呈现出明显的周期性变化规律。例如，机组负荷通常会按照日、周的周期出现波动，这种周期性是机组运行规律的重要体现。利用时间序列预测模型，能够充分考虑负荷数据的周期性特征，结合历史负荷数据的变化趋势，对缺失的负荷数据进行较为准确的补全。通过这种方式，可以为机组运行优化提供完整、可靠的数据支持，帮助运行人员更好地掌握机组运行状态，制定合理的运行策略，提高火电厂的运行效率和经济效益。

3.3 基于机器学习的补全策略

在热工测量数据缺失值补全领域，机器学习算法展现出强大的优势，随机森林、神经网络等算法成为热门选择。这些算法的核心能力在于能够深度挖掘数据间复杂的非线性关系，通过大量数据的训练，构建出精准的预测模型，进而实现对缺失值的可靠估算。以随机森林算法为例，在处理压力数据缺失问题时，可将与缺失压力数据相关的多个热工参数，如温度、流量等作为输入特征。算法通过对大量历史数据的学习，掌握这些特征与压力数据之间的内在联系，进而预测出缺失的压力值。相较于传统的缺失值补全方法，基于机器学习的补全策略在处理复杂数据关系时具有更高的补全精度，能够更准确地还原数据的真实情况。然而，该策略也存在一定局限性，模型训练需要依赖大量的历史数据，以确保模型具有足够的泛化能力。同时，由于算法本身的复杂性，其计算复杂度较高，对计算资源和时间成本有一定要求。

4 火电厂热工测量数据异常值处理方法

4.1 统计方法检测异常值

基于统计原理的异常值检测方法是数据处理中常用的手段，其中 3σ 原则（三倍标准差法则）是较为经典的

一种。该方法基于数据服从正态分布的假设,认为大部分数据会集中在均值附近,而超出均值三倍标准差范围的数据则被判定为异常值。这是因为,在正态分布中,数据落在均值加减三倍标准差范围内的概率极高,超出该范围的数据出现的概率极小,很可能是由异常因素导致的。在火电厂热工测量数据处理场景中,对于部分符合正态分布特征的参数, 3σ 原则能够发挥重要作用。例如,在机组稳定运行状态下,部分蒸汽压力数据通常呈现出正态分布特征。此时运用 3σ 原则,可以快速识别出那些明显偏离正常范围的异常压力值。这些异常值可能暗示着设备存在故障或运行异常,及时检测出它们能够为设备故障预警提供有力依据,帮助运行人员提前采取措施,保障火电厂的安全稳定运行^[4]。

4.2 基于距离的异常值检测

基于距离的异常值检测方法是一种从数据空间分布角度出发,识别异常数据的有效手段,局部离群因子(LOF)算法是其中的典型代表。该方法的核心思想是通过计算数据点与其他数据点之间的距离来评估其异常程度。与传统的仅依据全局距离判断异常的方法不同,LOF算法着重考虑数据点的局部密度。它能够敏锐地捕捉到数据分布不均匀的情况,对于那些在局部区域内显得孤立的数据点,即便其在全局范围内距离其他数据点不算远,也能被准确识别为异常值。在火电厂热工测量数据处理中,常常需要对多个热工参数进行联合分析。此时,数据处于多维数据空间中。LOF算法能够在这个复杂的多维空间里,精准地找出那些与周围数据点距离较远的异常数据组合。这些异常数据组合往往反映了设备运行过程中的异常状况,通过对它们的深入分析,可以帮助运行人员快速定位设备运行异常的原因,为设备的及时维护和故障排除提供有力支持。

4.3 深度学习异常值检测模型

深度学习模型为热工测量数据异常值检测提供了新的有效途径,自动编码器(AutoEncoder)便是其中颇具代表性的模型。自动编码器的工作原理是先对正常数据进行学习,捕捉其内在的特征表示,然后利用这些特征

对输入数据进行重构。在正常数据输入时,由于模型已经充分学习了正常数据的模式,重构误差相对较小;而当输入异常数据时,由于异常数据偏离了正常数据的特征模式,模型无法准确重构,重构误差会显著增大。通过设定合适的阈值,当重构误差超过该阈值时,即可判定数据为异常。在火电厂的实际应用场景中,以发电机振动监测数据为例,利用自动编码器对大量正常振动数据进行训练。在设备运行过程中,持续将实时采集的振动数据输入训练好的模型。一旦模型检测到振动数据的重构误差增大,就意味着振动数据出现了异常变化,这可能预示着发电机存在潜在故障。通过这种方式,能够提前发出预警,为设备维护人员争取宝贵的处理时间,有效提高设备的运行可靠性。

结束语

综上所述,火电厂热工测量数据具有实时性强、多源性、非线性和时变性等特点,且存在噪声、缺失值和异常值等问题,严重影响数据可用性。针对这些问题,本文详细介绍了多种处理方法。在噪声处理上,滤波算法各展所长;缺失值处理方面,插值、时间序列预测和机器学习补全策略各有适用场景;异常值检测中,统计方法、基于距离的方法和深度学习模型提供了不同维度的解决方案。通过合理运用这些数据预处理方法,能有效提升热工测量数据质量,为火电厂的数据分析、故障诊断、性能评估及控制决策提供可靠支撑,助力火电厂实现智能化、高效化与安全稳定运行。

参考文献

- [1]刘香云.火电厂热工测量的常见问题及改进措施[J].湖北电力,2020(2):30-32.
- [2]宋联伟.论火电厂热工测量的常见问题与解决[J].科技致富向导,2021(24):375.
- [3]薛辉,陆晔.火电厂热工参数软测量技术的应用价值分析[J].科学技术创新,2019(26):20-21.
- [4]王冬生.火电厂热工自动化中自动控制理论及实际应用研究[J].应用能源技术,2020(10):14-16.