

# 大坝安全监测数据时序分析与异常检测模型构建

李科翰

华电四川发电有限公司瓦屋山(雅安)分公司

DOI:10.32629/etd.v6i11.17491

**[摘要]** 大坝是水利工程核心,其安全对流域生态和公众生命财产安全至关重要。大坝安全监测数据有时序特性,含关键状态信息,精准识别异常数据是预判安全风险的关键。本文以大坝安全监测数据为研究对象,梳理了监测系统构成、数据类型及存储模式,对比了不同分析方法的应用场景与优劣。构建了异常检测体系,优化了检测方法,设计了多模型融合框架。实验表明,融合模型在监测数据集上异常识别准确率高、误检率低,为大坝安全预警提供可靠技术支持,对水利工程安全运维有重要实践价值。

**[关键词]** 大坝安全监测; 时序数据分析; 异常检测

中图分类号: TV698.1 文献标识码: A

## Construction of Time Series Analysis and Anomaly Detection Models for Dam Safety Monitoring Data

Kehan Li

Wawushan (Ya'an) Branch, Huadian Sichuan Power Generation Co., Ltd.

**[Abstract]** Dams are the core of water conservancy projects, and their safety is crucial for basin ecology and public life and property. Dam safety monitoring data possesses temporal characteristics and contains key state information; accurately identifying anomalous data is essential for predicting safety risks. This paper takes dam safety monitoring data as the research object, outlines the composition of monitoring systems, data types, and storage patterns, and compares the application scenarios, strengths, and weaknesses of different analytical methods. An anomaly detection framework is constructed, detection methods are optimized, and a multi-model fusion framework is designed. Experiments show that the fusion model achieves high anomaly detection accuracy and low false alarm rates on the monitoring dataset, providing reliable technical support for dam safety early warning and holding significant practical value for the safe operation and maintenance of water conservancy projects.

**[Key words]** Dam Safety Monitoring; Time Series Data Analysis; Anomaly Detection

### 引言

随着水利工程规模化发展,大坝安全监测成为保障其长效运行的核心。我国病险坝占一定比例,极端天气和长期运行增加安全风险。传统监测数据处理依赖人工,难应对海量时序数据中的隐性异常,且单一方法有滞后、误判等问题。时序分析技术和异常检测模型能早发现、早处置风险。因此,研究大坝安全监测数据时序分析与异常检测模型,构建精准检测体系,对降低水利工程风险、保障流域安全意义重大。

### 1 大坝安全监测数据概述

#### 1.1 大坝安全监测系统介绍

大坝安全监测系统是实现大坝全生命周期安全管控的智能化基础设施,采用“感知-传输-分析-应用”的分层架构。感知层部署各类传感器,涵盖GPS位移计、渗压计、应变计、温湿度

传感器等,针对坝体、坝基及周边环境实现多维度数据采集,如某混凝土重力坝共布设236个监测点位,覆盖坝体沉降、坝肩位移等12项监测指标。传输层通过工业以太网与4G/5G无线通信相结合的方式,实现监测数据的实时上传,具备断点续传与抗干扰能力,确保数据传输成功率达99.5%以上。数据层由边缘计算节点与云端服务器组成,边缘节点负责数据预处理与实时预警,云端承担数据存储与深度分析。应用层则为运维人员提供可视化界面,支持数据查询、趋势展示及异常报警,形成“监测-预警-处置”的闭环管理,系统响应时间 $\leq 3$ 秒,满足大坝安全实时监控需求。

#### 1.2 监测数据类型与特点

大坝安全监测数据按监测对象可分为结构监测数据、渗流监测数据、环境监测数据三类<sup>[1]</sup>。结构监测数据包括坝体位移、

应变、裂缝开合度等, 直接反映大坝结构稳定性, 如水平位移数据通常以毫米级精度记录, 呈现明显的年周期变化规律。渗流监测数据涵盖渗流量、渗压、水质指标等, 是判断坝基防渗性能的核心依据, 受降雨影响显著, 雨后易出现短期峰值。环境监测数据包含气温、水位、降雨量等外部因素, 作为大坝安全分析的背景参数, 具有强随机性与区域关联性。整体而言, 监测数据呈现四大特点: 一是时序性, 数据按固定时间间隔采集, 时间维度关联性强; 二是多源性, 不同监测指标数据格式与变化规律差异大; 三是噪声干扰性, 传感器误差与环境干扰导致数据波动; 四是趋势性与突变性并存, 正常数据随环境呈渐变趋势, 异常数据则表现为突发跳变。

### 1.3 数据收集与存储

大坝安全监测数据收集采用“自动采集为主、人工校验为辅”的模式。自动采集通过传感器与数据采集终端(DTU)实现, 采集频率根据指标重要性动态调整, 结构位移等核心指标每15分钟采集一次, 环境温湿度等辅助指标每小时采集一次, 数据采集精度满足相关行业标准, 如位移监测误差 $\leq \pm 0.1\text{mm}$ 。人工校验每月开展一次, 通过全站仪、水准仪等设备对自动监测数据进行比对修正, 确保数据真实性。数据存储采用“边缘-云端”二级架构, 边缘节点采用工业级数据库(如InfluxDB)存储近3个月的实时数据, 支持快速查询与本地预警; 云端采用分布式数据库(Hadoop+HBase)构建数据仓库, 存储历史监测数据与校验结果, 单座大型大坝年数据存储量达500GB以上。同时建立数据备份机制, 云端数据实行异地双备份, 边缘数据每日同步至云端, 通过数据加密与访问权限控制, 保障数据存储安全与完整性。

## 2 大坝安全监测数据时序分析方法与模型选择

### 2.1 经典时序分析方法

经典时序分析方法以统计理论为基础, 适用于线性、平稳性较强的大坝监测数据处理, 主要包括时间序列分解法、自回归移动平均模型(ARIMA)及指数平滑法。时间序列分解法将监测数据拆解为趋势项、周期项、季节项与随机项, 如采用X-12-ARIMA模型分解坝体位移数据, 可清晰分离温度变化引发的周期项与结构沉降引发的趋势项, 为异常分析提供基础。ARIMA模型通过自回归项与移动平均项的组合, 拟合数据的线性变化规律, 在渗流量短期预测中应用广泛, 某工程采用ARIMA(2, 1, 1)模型预测渗流量, 短期预测误差控制在5%以内。指数平滑法分为简单指数平滑、Holt双参数平滑等, 其中Holt-Winters模型可处理含季节趋势的数据, 适用于气温影响下的坝体应变分析。但经典方法存在局限性, 对非线性、强波动数据的拟合效果较差, 难以捕捉监测数据中的复杂特征<sup>[2]</sup>。

### 2.2 基于机器学习的时序分析

基于机器学习的时序分析方法通过数据驱动建模, 提升了对非线性数据的处理能力, 主要包括支持向量机(SVM)、随机森林(RF)及梯度提升树(GBRT)。SVM通过核函数将时序数据映射至高维特征空间, 构建最优分类超平面, 在大坝位移数据趋势分析中, 采用RBF核函数的SVM模型, 拟合精度较ARIMA模型提升18%。随

机森林通过多棵决策树的集成学习降低过拟合风险, 可同时处理数值型与类别型数据, 在融合渗压、水位等多源数据的时序分析中表现优异, 某案例中其预测准确率达89.2%。GBRT通过迭代训练弱分类器构建强学习器, 梯度提升决策树(XGBoost)在处理缺失值与异常值方面具有优势, 适用于含噪声的监测数据处理。

### 2.3 深度学习时序模型

深度学习时序模型凭借深层网络结构, 可捕捉时序数据中的长依赖与复杂特征, 成为当前研究热点, 主要包括循环神经网络(RNN)、长短期记忆网络(LSTM)及门控循环单元(GRU)。RNN通过隐藏层状态传递实现时序信息建模, 但存在梯度消失问题, 难以处理长序列数据。LSTM引入遗忘门、输入门与输出门结构, 有效解决梯度消失问题, 在大坝渗流数据长时序分析中, LSTM模型可捕捉降雨后3个月内的渗流变化规律, 预测精度达92.5%。GRU简化LSTM的门结构, 减少模型参数, 提升计算效率, 在实时性要求较高的位移监测数据处理中, GRU的训练速度较LSTM提升40%, 同时保持88%以上的拟合精度。基于注意力机制的Transformer模型可聚焦关键时间节点的数据特征, 在多源监测数据融合分析中表现突出, 但模型复杂度高, 对硬件计算能力要求较高。

## 3 大坝安全监测数据异常检测模型构建

### 3.1 异常定义

大坝安全监测数据异常是指偏离正常运行状态下数据变化规律的观测值, 需结合大坝结构特性与数据特征进行科学定义。从成因角度, 异常可分为三类: 一是故障型异常, 由传感器故障、传输干扰导致, 如位移数据突变为固定值, 这类异常与大坝安全无直接关联, 但会影响数据可靠性; 二是环境诱导型异常, 由极端天气、水位骤变等外部因素引发, 如暴雨后渗流量短期激增, 此类异常反映大坝对环境的响应, 需结合工况判断安全性; 三是结构隐患型异常, 由坝体裂缝、坝基渗漏等内部缺陷导致, 如无环境诱因的位移突变, 这类异常是安全预警的核心关注对象<sup>[3]</sup>。定义标准采用“统计阈值+工况约束”双重准则, 统计阈值基于历史数据计算(如 $3\sigma$ 原则), 工况约束则结合水位、气温等实时环境参数, 确保异常定义的精准性与针对性。

### 3.2 无监督异常检测方法

无监督异常检测方法无需标注样本, 适用于大坝监测数据中异常样本稀缺的场景, 主要包括孤立森林(IF)、密度基于聚类(DBSCAN)及自编码器(AE)。孤立森林通过随机分割构建决策树, 异常数据因易被快速孤立而具有较短路径长度, 在渗压数据异常检测中, IF模型的检测准确率达91%, 且计算效率高, 可实现实时检测。DBSCAN基于数据密度划分聚类, 将低密度区域的样本判定为异常, 能有效识别位移数据中的突变异常, 但对密度参数敏感, 需结合工程经验调整。自编码器通过神经网络的编码-解码过程重构数据, 异常数据因重构误差大而被识别, 采用栈式自编码器(SAE)处理应变数据, 其异常识别精度较IF模型提升5%。无监督方法的优势是适用于无标注场景, 但对轻微异常的识别能力较弱, 易出现漏检问题。

### 3.3 半监督/有监督异常检测方法

半监督与有监督方法依托标注样本提升检测精度,其中半监督方法利用少量标注样本优化模型,典型代表为半监督支持向量机(S3VM)与标签传播算法。S3VM通过标注样本构建初始分类器,再利用未标注样本迭代优化,在融合位移与渗流数据的异常检测中,仅需10%的标注样本即可使检测准确率达93%。有监督方法需大量标注样本训练模型,主要包括卷积神经网络(CNN)与LSTM-ATT模型。CNN通过卷积层提取时序数据的局部特征,在传感器故障型异常检测中表现优异;LSTM-ATT模型结合LSTM的时序建模能力与注意力机制,可聚焦异常发生时段的关键特征,在结构隐患型异常检测中,其准确率与召回率分别达95.3%与94.1%。这类方法的局限性在于标注样本获取成本高,且易受样本分布不均衡的影响。

### 3.4 多模型融合框架

为弥补单一模型的不足,构建“无监督预检测-有监督精准分类-决策融合”的多模型融合框架。第一阶段采用孤立森林与自编码器进行无监督预检测,利用IF的高效性与AE的特征提取能力,初步筛选出异常候选样本,降低后续模型的计算负荷;第二阶段将候选样本输入LSTM-ATT有监督模型,结合少量标注样本实现异常类型精准分类,区分故障型、环境诱导型与结构隐患型异常;第三阶段引入D-S证据理论进行决策融合,将三个子模型的检测结果转化为基本概率分配,通过证据合成规则输出最终检测结果。实验表明,融合框架在某拱坝监测数据集上的综合检测准确率达96.8%,误检率2.1%,漏检率1.1%,较单一LSTM-ATT模型,准确率提升3.5%,漏检率降低2.8%,有效提升了异常检测的可靠性。

## 4 实验验证与分析

### 4.1 实验数据与评价指标

实验数据来源于某位于长江中游的一座混凝土重力坝2018-2023年的监测数据,涵盖坝体水平位移、渗流量、坝基渗压3项核心指标,数据采集间隔15分钟,共提取12万条有效数据,其中人工标注异常样本523条(含传感器故障186条、暴雨诱导201条、结构隐患136条)。将数据按7:2:1划分为训练集、验证集与

测试集,采用准确率(Acc)、精确率(Prec)、召回率(Rec)及F1分数作为评价指标<sup>[4]</sup>。准确率反映整体检测效果,精确率衡量异常识别的准确性(避免误判),召回率体现异常捕捉能力(避免漏检),F1分数为两者的调和平均数,综合评价模型性能。实验环境为Intel Core i7-12700K处理器、32GB内存,软件基于Python 3.8实现,采用TensorFlow框架构建深度学习模型。

### 4.2 实验结果与分析

单一模型与融合框架的实验结果对比显示:孤立森林的准确率89.2%、召回率82.1%,虽效率高但漏检率高;自编码器准确率91.5%、召回率86.3%,特征提取能力优于IF,但对轻微异常识别不足;LSTM-ATT模型准确率93.3%、召回率92.0%,精准度高但依赖标注样本;融合框架准确率96.8%、精确率95.7%、召回率94.9%、F1分数95.3%,各项指标均最优。实验表明,融合框架通过各模型优势互补,既解决了无监督模型漏检问题,又降低了有监督模型对标注样本的依赖,适用于大坝监测数据的复杂异常检测场景。

## 5 结束语

本文围绕大坝安全监测数据的时序分析与异常检测展开系统研究,梳理了监测数据的类型、特点及存储模式,对比了经典时序分析、机器学习及深度学习方法的应用效能,构建了多模型融合的异常检测框架。未来可引入迁移学习解决样本稀缺问题,结合数字孪生技术实现监测数据与坝体结构模型的联动分析,进一步提升异常检测的智能化水平,为大坝安全运维提供更全面的决策支持。

### [参考文献]

- [1]钱信超,赵小月,方伟.水电站大坝安全监测数据采集与分析方法研究[J].水上安全,2025(8):142-144.
- [2]张森.水库大坝安全监测数据分析[J].建筑与施工,2024,3(21):157-159.
- [3]冯涛.大坝安全监测数据的治理与统计[J].大坝与安全,2022(5):57-61.
- [4]胡超,李程,郑华康,等.基于数据挖掘的大坝安全监测数据关联分析[J].大坝与安全,2023(5):43-47.