

# 基于深层神经网络的长大桥梁监测数据异常修复与重构方法研究

王连发<sup>1,2,4</sup> 徐一超<sup>1,2,4</sup> 王昆鹏<sup>3</sup> 张宇峰<sup>1,2,4</sup> 承宇<sup>1,2,4</sup>

1. 长大桥梁安全长寿与健康运维全国重点实验室 南京 210019

2. 苏交科集团股份有限公司 南京 211199

3. 中交公路长大桥梁建设国家工程研究中心有限公司 北京 100088

4. 广东粤交科科技有限公司 广州 511462

**摘要:** 针对长大桥梁结构健康监测中因传感器故障、环境干扰等导致的数据异常与缺失问题, 本研究提出了一种基于深层神经网络的数据修复与重构方法。该方法融合双向长短期记忆网络 (BiLSTM) 和条件生成对抗网络 (CGAN), 充分挖掘多源异构监测数据的时空关联, 实现异常特征的精准捕捉与缺失序列的高精度重构。实验表明, 该方法有效提升了数据质量, 平均绝对误差 (MAE) 降低 21.8%, 均方根误差 (RMSE) 降低 42.7%, 模型拟合能力提高 9.1%。

**关键词:** 深层神经网络; 桥梁健康监测; 数据修复; 异常重构

## 引言

长大桥梁是交通基础设施的重要组成部分, 因此长大桥梁结构的安全保障显得极为重要。结构健康监测 (SHM) 系统就是一种针对桥梁长期监测的平台, 能够实时采集到桥体结构的各种振动、应变、温度等多维信息, 是对桥梁进行健康状态评估的重要手段。而基于实际工程的运用, 存在着传感器老化故障、环境噪声干扰、传输丢包及压缩失真等诸多问题因素导致监测数据出现诸如缺少、跳点、噪声、漂移、趋势偏离等数据问题。国内外学者将桥梁结构健康监测数据的异常种类主要分为缺失、噪声、增益、精度下降、漂移、趋势、次小值、离群值和超量程振荡等 9 类<sup>[1]</sup>。工程界前期开展了 10 余年海量桥梁结构监测数据处理工作, 归纳总结出长大桥梁结构监测主要存在的数据异常包括: 数据缺失、数据跃迁、长周干扰、非一致异点 (跳点)、弱噪声干扰、无法消除的强噪声、其他难以快速识别和消除的数据异常等 7 类<sup>[2]</sup>。这些分离基本涵盖了目前结构监测系统数据异常的有所形式。

本文针对以上核心问题, 开展基于深层神经网络的数据修复与重构方法研究。通过融合双向长短期记忆网络 (BiLSTM) 和条件生成对抗网络 (CGAN), 充分挖掘多源异构监测数据的时空关联, 实现异常特征的精准捕捉与缺失序列的高精度重构。

## 1 数据异常问题

### 1.1 监测数据异常类型

桥梁健康监测数据异常主要表现为以下几种类型:

数据缺失: 因为传感器异常或者数据传输问题等导致部分或者全部数据丢失, 根据其缺失模式的不同可以将其分为随机缺失、连续缺失及整个通道缺失等类型。

(1) 跳点异常: 数据中出现瞬时尖峰或者脉冲, 其原因是外部电磁干扰和采集系统瞬时故障所致。

(2) 噪声异常: 由于存在着环境干扰、测量误差以及电子设备热噪声等因素, 所以数据中必然存在一定的随机波动。

(3) 信号漂移: 由于传感器基准值随时间产生一种趋势性的偏移, 一般情况下, 这是由传感器老化和温度变化所引起的。

(4) 趋势异常: 数据长时间处于和正常情况不一样的模式下运行, 有可能是结构受损、或者环境发生了改变。

表 1: 桥梁监测数据异常类型及特征

异常类型	主要特征	产生原因	处理难度
数据缺失	数据点或段完全丢失	传感器故障、传输中断	中等
跳点异常	瞬时尖峰脉冲	电磁干扰、瞬时故障	低
噪声异常	随机波动	环境干扰、测量误差	低-中等
信号漂移	基准值缓慢变化	传感器老化、温度变化	高
趋势异常	长期趋势变化	结构损伤、环境变化	极高

### 1.2 传统处理方法的局限性

传统的数据异常处理方法主要有通过可视化方式查看、

使用数据统计学检验及通过模型分析等方法找出异常值，并用删除或者替换等方法对其进行处理，但这些传统方法有明显的局限性 [3,4]。更重要的是，目前很多方法忽略了桥梁结构本身所具备的力学特性，并没有把结构物理先验知识应用到桥梁的状态评价当中去，所以修复后的结果难免存在脱离结构力学实际的可能，会影响到后期状态评估的正确性。

## 2 方法理论

本研究提出的基于深层神经网络的数据修复与重构方法整体框架主要包括数据预处理、特征提取、异常诊断与数据修复四个核心模块。

### 2.1 数据预处理与特征提取

数据预处理能够保证数据分析结果的准确性。首先要对原始监测数据进行清洗，然后运用基于统计分布的方法找寻并剔除明显的异常值；对于缺失数据，在判断其缺失比例之后采取不同的措施，当缺失比例不大时，利用均值、中位数或者众数填补空白；若是缺失比例大，则根据数据自身时间的相关性，采用插值法填补或者采用某一相近时期的数据来修补。

考虑到不同传感器采集的数据量纲和数值范围差异较大，直接分析会影响结果准确性，需对数据进行归一化处理。采用离差标准化方法将数据映射到 [0,1] 区间，公式如下：

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

其中， $x_{\min}$  和  $x_{\max}$  分别为数据的最小值和最大值。

按照以上要求，在每分钟内获得归一化数据时生成 11 维特征向量，分别为：均值、中值、标准差、均方根、幅值、差值、百分二十位数、百分八十位数、最大值与百分八十位数之比、均值与幅度之比以及均值与百分八十位数之比。该特征是从 40 个统计特征指标库中经过多轮对比筛选后所得出的较为敏感的特征点，对模型而言是较为重要的部分，能在一定程度上提高模型的准确度，减少训练成本 [5]。

### 2.2 神经网络模型设计

#### 2.2.1 双向长短期记忆网络 (BiLSTM) 异常诊断

为了解决监测数据异常诊断的问题，采用包含三个双向 LSTM 层、一层全连接层和一层 softmax 层的 BiLSTM 模型。其中，基于 BiLSTM 模型可以同时学习时序数据的前向和后向信息的特点，来获取监测数据中蕴含的时序特征。

#### 2.2.2 条件生成对抗网络 (CGAN) 数据修复

为了解决异常数据修复问题，利用 CGAN 进行数据重建。CGAN 是利用添加了条件变量，使生成器、判别器得以通过数据指导生成过程，完成多传感器间非线性关系学习的数据重建过程。

生成器 G 试图学习真实数据分布，以随机噪声  $z$  和条件向量  $c$  为输入，生成修复数据  $G(z|c)$ ；判别器 D 则尝试区分真实数据  $x$  和生成数据  $G(z|c)$ ，同时考虑条件向量  $c$ 。CGAN 的价值函数可表示为：

$$\begin{aligned} & \min_G \max_D V(D, G) \\ & = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x|c)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z|c)|c))] \end{aligned}$$

经过对抗训练，生成器能够学习产生难以区分的真实数据分布，实现异常数据的精修复。

#### 2.2.3 深度神经辐射场精细化重建

为了满足桥梁故障检测以及精细化重建的需求，运用深度神经辐射场 (NeRF)，仅依靠少量图片即可还原出高分辨率、高频细节的 3D 场景，适用于复现复杂环境中多角度几何细节；以合成的 3D 高分辨率深度场与桥梁标准深度图作为参照，并进行对比，可精准地检测到微小损伤部位 [6]。

### 2.3 多源时空数据融合

为了充分发挥监测数据的时空特性优势，提出一种多源时空数据融合方法：以卷积神经网络模块抽取数据时序局部特征，利用通道注意力模块动态量测点间空间相关性，使用双向长短期记忆模块结合融合后的特征开展序列分析建模。此多源融合方法同时考虑了监测数据的时间依存性和空间相关性，充分发挥了多源传感器数据之间的内在联系性，提高了数据修复重构的精度；可以根据不同桥梁监测指标的不同特点自适应调整融合权重，达到个性化的目的 [7]。

## 3 实验验证与结果分析

### 3.1 实验设置与评价指标

为验证所提出方法的有效性，分别进行了数值模拟和实桥监测数据实验。采用某大跨度斜拉桥的 GPS 监测数据作为实验数据集，该数据集包含 8 个传感器采集的结构响应数据。

采用以下评价指标量化方法性能：

平均绝对误差 (MAE)：衡量修复数据与真实数据之间的平均绝对差异

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

均方根误差 (RMSE)：衡量修复数据与真实数据之间的平均平方根差异

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

拟合能力 (R<sup>2</sup>)：衡量修复数据对真实数据的解释程度

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

### 3.2 结果分析

对数值模拟实验中人为造成一定的数值缺损或错误值的比例不同,采用传统方法与本文所提的神经网络方法进行数据修复、重构,得到的结果见表 2。

表 2: 不同数据修复方法的性能比较

数据缺失比例	方法	MAE	RMSE	R <sup>2</sup>	计算时间 (s)
10%	线性插值	0.124	0.156	0.872	0.5
10%	ARIMA	0.098	0.132	0.901	2.3
10%	本文方法	0.067	0.089	0.958	3.8
30%	线性插值	0.253	0.312	0.723	0.6
30%	ARIMA	0.201	0.254	0.785	2.5
30%	本文方法	0.112	0.145	0.912	4.2
50%	线性插值	0.412	0.498	0.532	0.6
50%	ARIMA	0.325	0.401	0.643	2.7
50%	本文方法	0.186	0.237	0.862	4.5

从表中可以看出,在各种方法的性能都随数据缺失比例的增大而不断减小的情况下,本文采用的神经网络法在整个过程中的效果都要好于其他的方法。其中,在数据缺失比例达到 50% 的时候,本文方法相较于传统方法,MAE、RMSE 与拟合能力分别减少 42.7%、40.9% 与 34.0%,同时由于增加了数据的量,有效增强了学习的训练量;但是它的计算时间相比传统方法会稍多一点,但是仍能符合工程实践中对计算效率的要求。

### 4 应用前景与挑战

基于深层神经网络的桥梁监测数据异常修复与重构方法在实际工程中具有广阔的应用前景:可以在当前桥梁健康监测系统中集成应用,在采集、处理和分析阶段做好相关数据的校验与修正工作,针对结构的状态信息实现高效合理的数据采集、在线判断及快速更新。优化后的质量监测数据可用于精准的结构状态诊断(损伤识别),安全预警和剩余寿命预测等工作,有利于科学化地指导桥梁养护管理<sup>[8]</sup>。

然而,这种方法还在使用上有许多困难之处。

计算资源需求:本文网络模型训练和推理所需要大量的计算资源对于嵌入式设备而言很难做到即时的、稳定的运行。后续可以采取模型轻量化、知识蒸馏等方式减小模型的计算量来解决该问题。

模型可解释性:深度学习模型是黑箱模型,难以理解其内部机制,也不易被人所信任,后续工作中应该加强引入可解释人工智能(XAI)来解释模型内部的机理<sup>[9]</sup>。

### 5 结论

本文针对长大桥梁结构健康监测系统常见的数据缺失、跳点、噪声、漂移及趋势异常等问题,构建了一套完整的深度学习体系。该方法融合了双向长短期记忆网络(BiLSTM)与条件生成对抗网络(CGAN),能够有效挖掘监测数据中的时空依赖关系,实现对异常数据的精准识别与高质量重构。

### 参考文献:

- [1] Liu G, Li L, Zhang L, et al. Sensor faults classification for SHM systems using deep learning-based method with Tsfresh features[J]. Smart Materials and Structures, 2020, 29(7):075005.
- [2] 张宇峰,李贤琪.桥梁结构健康监测与状态评估[M].上海科学技术出版社,2018.
- [3] Pimentel M A, Clifton D A, Clifton L, et al. A review of novelty detection [J]. Signal Processing, 2014, 99: 215–49.
- [4] 唐志一.基于深度学习的结构健康监测异常数据诊断与重构[D].哈尔滨工业大学,2021.
- [5] 徐一超,张宇峰,承宇,张建.面向场景需求的桥梁轻量化监测技术与装备[C]//中国振动工程学会结构抗振控制与健康监测专业委员会.第九届全国结构抗振控制与健康监测学术会议摘要集.[出版者不详],2023:221.
- [6] 长大桥梁结构监测数据挖掘分析[J].中国公路,2023(10):48–51.
- [7] 马正源.基于结构监测的长大桥梁工程运营期风险管理研究[D].北京邮电大学:2024.D
- [8] 李志华,刘子振,邵国涛,何春木.深厚软土区长大桥梁变形趋势监测研究[J].交通世界,2024(18):155–158.
- [9] 《公路长大桥梁结构监测时空大数据应用指引》解读[J].城市道桥与防洪,2024(11):309–310.