

DOI:10.19651/j.cnki.emt.2518697

基于多级注意力机制的滑坡位移多步预测方法^{*}

任 冯 肖 慧 冯 沂 萱 吴 雨 洁 艾 玉 洁

(东华理工大学地球物理与测控技术学院 南昌 330013)

摘要: 针对土质滑坡位移多步预测方法的缺乏以及在多时间步长下预测误差较大的问题,本文提出了一种基于多级注意力机制并行模型的滑坡位移多步预测方法。采用多输入多输出的预测策略,通过含有多头注意力机制的Transformer编码器网络分支以及经全局注意力机制(GAM)优化的双向门控循环单元(BiGRU)网络分支,两个网络分支并行处理滑坡历史监测数据,之后对并行网络提取到的滑坡特征信息通过交叉注意力机制(CAM)进行特征融合后输出预测的滑坡多步位移值。实验结果表明,多级注意力机制模型在滑坡位移多步预测中平均绝对误差(MAE)、均方根误差(RMSE)分别为2.17 mm、3.05 mm,决定系数(R^2)为0.968 9,相较于其他模型误差最低,决定系数结果最优,在长时间步下的预测效果更加稳定,有利于提前预知滑坡发展动向,为滑坡的预防与治理提供了重要的技术支持。

关键词: 滑坡位移多步预测;多级注意力机制;Transformer编码器;双向门控循环单元;多输入多输出策略

中图分类号: TN306;TP18;P642.22 **文献标识码:** A **国家标准学科分类代码:** 620.1030

Multi-step prediction method for landslide displacement based on multi-level attention mechanism

Ren Feng Xiao Hui Feng Yixuan Wu Yujie Ai Yujie

(School of Geophysics and Measurement-Control Technology, East China University of Technology, Nanchang 330013, China)

Abstract: Aiming at the lack of multi-step prediction methods for soil landslide displacement and the issue of significant prediction errors over extended time horizons, this paper proposes a multi-step landslide displacement prediction method based on a parallel model with a multi-level attention mechanism. The method employs a multi-input multi-output prediction strategy. Utilizing a Transformer encoder branch incorporating a multi-head attention mechanism and a bidirectional gated recurrent unit (BiGRU) branch optimized with a global attention mechanism (GAM), the two parallel network branches process historical landslide monitoring data. The landslide feature information extracted by the parallel networks is then fused via a cross attention mechanism (CAM), subsequently outputting the predicted multi-step displacement values. Experimental results demonstrate that the multi-level attention mechanism model achieves a mean absolute error (MAE) of 2.17 mm, a root mean square error (RMSE) of 3.05 mm, and a coefficient of determination (R^2) of 0.968 9 in multi-step landslide displacement prediction. Compared to other models, it yields the lowest errors and the optimal R^2 result. The model exhibits more stable prediction performance over long time horizons, facilitating the early anticipation of landslide development trends. This provides crucial technical support for landslide prevention and mitigation.

Keywords: landslide displacement prediction; multi-level attention mechanisms; Transformer encoder; bidirectional gated recurrent unit; multi-input multi-output strategy

0 引 言

土质滑坡位移受土壤内部性质和外界环境的影响表现为长时间的波动变化,这种波动性导致滑坡位移预测的时

效性不长,多以单步预测为主。单步预测仅能预测后一步的滑坡位移值,无法提供更长的预警反应时间^[1]。目前滑坡位移多步预测方法在预测时间步长和准确性上的效果同样不佳^[2],难以提前掌握滑坡发展动向。为了提供更长的

收稿日期:2025-04-27

^{*} 基金项目:江西省自然科学基金(20212BAB203004)、东华理工大学研究生创新基金(YC2023-S579)项目资助

预警决策时间,适应滑坡早期预警和实际工程的应用,需要探索一种准确有效的滑坡位移多步预测方法。

滑坡多步预测方法的适用性不仅受滑坡数据非线性波动的影响,同样受多步预测策略选择的制约^[3]。对于具有长时间非线性特征的滑坡位移数据,选择合适的多步预测策略可以有效提高滑坡位移多步预测的精度。目前已有的研究中,滑坡位移预测策略通常以直接预测和滚动预测的预测策略为主,这两种预测策略会造成模型复杂度较高和误差累积的问题。雷德鑫等^[4]基于 ARIMA 时间序列模型实现了对滑坡累积位移的滚动预测。汪标等^[5]结合滚动预测的思想,预测了滑坡的周期性位移和随机性位移预测,结果能较好的符合位移整体变化趋势。但上述方法主要以单步预测或滚动预测为主,在单步或短期的预测中表现优秀,随着预测步长的增加,由于误差累积的原因,预测误差会逐渐增大。为了提高多步预测的准确度,采用多输入多输出的预测策略,将输入输出集合为一个整体,避免出现误差累积的问题,实现滑坡位移的多步预测。

采用适合处理滑坡非线性数据的模型与算法能够显著提升滑坡多步预测的准确度。随着深度学习等模型和算法的发展,相关的模型方法已广泛运用于滑坡位移预测领域,尤其是门控循环单元 (gate recurrent unit, GRU)、Transformer 模型等^[6-8]深度学习模型在滑坡位移预测方面发挥了重要的作用。罗祎沉等^[9]在滑坡位移时序的基础上,结合 GRU 实现对新型滑坡位移综合预测。Ge 等^[10]通过对 Transformer 模型的改进实现了对滑坡位移的精准预测;田原等^[11]引入 Transformer 模型构建了滑坡位移短期预测模型,在未来 3 个时间步长预测中取得了较好的效果。但 GRU 模型作为一种改进型的循环神经网络 (recurrent neural network, RNN),在缓解梯度消失和梯度爆炸等问题的同时又缺乏对全局特征信息的捕捉能力^[12-13]。Transformer 模型能够考虑滑坡数据中不同位置之间的影响关系,提取滑坡数据中全局的特征信息,在多个领域已得到广泛应用,但在长时间序列的处理上存在性能下降和计算瓶颈等问题。

针对上述问题,本文提出一种基于多级注意力机制并行模型的滑坡位移多步预测方法,用于实现长时间步下的滑坡位移多步预测。该并行模型分为两个分支,一是含有多头注意力机制的 Transformer 编码器网络分支;二是通过全局注意力机制 (global attention mechanism, GAM) 优化后的双向门控循环单元 (bidirectional gated recurrent unit, BiGRU) 网络分支,两个分支并行处理滑坡监测数据,充分捕获滑坡数据中长期的依赖关系和全局特征,之后通过交叉注意力机制 (cross attention mechanism, CAM) 对两分支提取到的特征信息进行融合。通过多级注意力机制并行模型的处理,能够同时预测多个时间步的滑坡位移值,在滑坡多步预测任务中拥有更好的性能和效果,提高了预测长时间步长的精度和效率。

1 多级注意力机制并行模型的构建

1.1 预测框架搭建

土质滑坡位移是具有时间关联性的数据,预测未来的滑坡位移值需要依据滑坡历史的监测数据进行处理、训练与预测。本文构建一种基于多级注意力机制并行模型的滑坡位移多步预测方法。

1) 首先对采集到的土质滑坡数据进行去噪、筛选、归一化等数据预处理,数据选用多输入多输出的预测策略,以满足滑坡数据的可靠性和适用性。

2) 之后采用两种分支方法进行并行处理,使用 Transformer 编码器网络分支和 BiGRU-GAM 分支。在 Transformer 编码器网络分支中,通过裁剪掉 Transformer 的解码器部分,只使用编码器进行数据关系捕捉与特征提取,降低模型的复杂度,优化其在长时间序列中的性能下降;在 BiGRU-GAM 分支中,使用 GAM 优化 BiGRU,通过上下文向量调整 BiGRU 对全局重点数据的捕捉能力,更有利于挖掘滑坡数据中上下依赖关系,提高对重要信息的感知能力。

3) 最后将两种分支方法得到的滑坡特征信息通过 CAM 进行特征融合,捕捉序列数据中的重要影响特征,整合输出得到滑坡位移的多步预测值,位移预测流程图如图 1 所示。

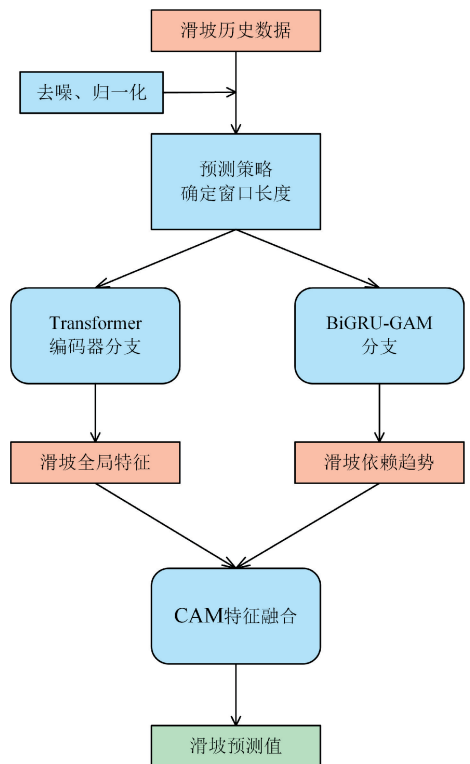


图 1 位移预测流程

Fig. 1 Cumulative displacement prediction process

1.2 多步预测策略的选取

为了避免滑坡位移滚动预测带来的误差累积的问题,本文采用多输入多输出的预测策略,使用滑动窗口的形式将滑坡的输入输出特征集合为单个样本,分别使用 12、20、28、36 的窗口长度来实现预测输出 3、5、7、9 步的滑坡位移值。

以 12 步长输入和 3 步长预测输出为例,即通过过去 12 步长的滑坡历史监测数据来预测捕捉未来 3 步的滑坡位移值,将滑坡输入输出数据通过定长的窗口集合起来,通过滑动窗口的形式向后滑动预测输出。每次预测该滑动窗口均向后移动,以此实现 3 步预测。同理,增加窗口的长度即可实现更长时间步的预测输出,预测策略示意图如图 2 所示。

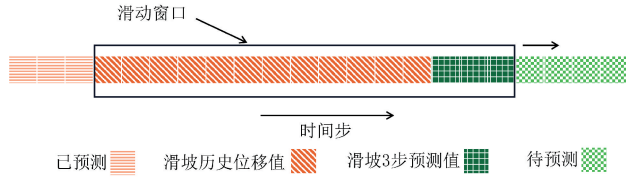


图 2 预测策略示意图

Fig. 2 Prediction strategy schematic diagram

1.3 Transformer 编码器分支

土质滑坡位移等特征数据经过位置编码后输出具有位置信息和滑坡特征信息的序列数据,作为 Transformer 编码器的输入。

Transformer 模型是一种基于注意力机制的编码器—解码器结构的深度学习模型^[14-15]。在本文中为了适应并行组合模型、降低模型复杂度,裁剪掉 Transformer 解码器部分,只使用 Transformer 编码器提取滑坡数据中的全局特征,加速了模型的训练过程以及为后续并行模型的特征输入提供特征数据基础。Transformer 编码器结构由多头注意力层、前馈神经网络层和两个残差归一化层构成。

滑坡特征数据在多头注意力层中计算不同序列位置之间影响关系的权重大小,通过注意力分数的形式给出,在前馈神经网络、残差归一化层作用下整合输出滑坡特征信息,在如式(1)~(5)所示。

$$\begin{cases} \mathbf{Q} = \mathbf{A} \times \mathbf{W}_Q \\ \mathbf{K} = \mathbf{A} \times \mathbf{W}_K \\ \mathbf{V} = \mathbf{A} \times \mathbf{W}_V \end{cases} \quad (1)$$

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_K}}\right)\mathbf{V} \quad (2)$$

$$head_h = Attention(\mathbf{Q}\mathbf{W}_{Q,h}, \mathbf{K}\mathbf{W}_{K,h}, \mathbf{V}\mathbf{W}_{V,h}) \quad (3)$$

$$Multi(\mathbf{A}) = concat(head_1, head_2, \dots, head_h)\mathbf{W}_{mul} \quad (4)$$

$$\mathbf{F} = \max(0, \mathbf{A}_1\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2 \quad (5)$$

式中: \mathbf{A} 、 \mathbf{Q} 、 \mathbf{K} 、 \mathbf{V} 表示输入的滑坡特征数据, $Attention(\cdot)$ 表示注意力分数, $head_h$ 为第 h 个注意力头,即计算的第 h 个注意力分数的结果。 $concat(\cdot)$ 表示拼接矩阵函数, $Multi(\mathbf{A})$ 表示多头的注意力分数, \mathbf{W} 表示各项矩阵参

数, \mathbf{b} 表示各项偏置项, $max(\cdot)$ 表示 ReLU 函数, \mathbf{F} 表示输出的滑坡全局特征数据。综上,Transformer 编码器示意图如图 3 所示。

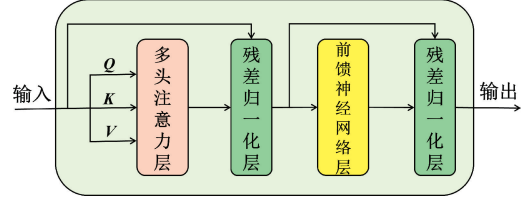


图 3 Transformer 编码器示意图

Fig. 3 Transformer encoder schematic diagram

1.4 BiGRU-GAM 分支

BiGRU 的基本单位是门控循环单元,该单元通过重置门和更新门可以实现对滑坡位移数据长短期局部特征的记忆与更新。针对 BiGRU 网络在全局特征捕捉方面的不足和对复杂的非线性数据处理上的劣势,本文应用 GAM 对 BiGRU 的输出进行加权处理,通过引入 GAM 提高对全局重要数据信息的捕捉与处理,使模型能够聚焦于滑坡位移序列中最重要的部分,提高模型的预测性能。

1) BiGRU

BiGRU 通过重置门和更新门对滑坡时间序列数据进行选择和更新,通过正反双向 GRU 捕获滑坡位移数据中的依赖趋势特征,当前时间步的滑坡特征信息输出如式(6)~(9)所示。

$$\mathbf{r}_t = \sigma([\mathbf{h}_{t-1}, \mathbf{x}_t]\mathbf{W}_r + \mathbf{b}_r) \quad (6)$$

$$\mathbf{z}_t = \sigma([\mathbf{h}_{t-1}, \mathbf{x}_t]\mathbf{W}_z + \mathbf{b}_z) \quad (7)$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}[\mathbf{r}_t \cdot \mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}) \quad (8)$$

$$\mathbf{h}_t = \mathbf{z}_t \cdot \tilde{\mathbf{h}}_t + (1 - \mathbf{z}_t) \cdot \mathbf{h}_{t-1} \quad (9)$$

式中: \mathbf{x}_t 表示该时间步输入的滑坡特征数据, \mathbf{h}_{t-1} 表示上一个时间步输出的滑坡特征信息, \mathbf{W} 表示各项参数矩阵, \mathbf{b} 表示各项偏置, $\sigma(\cdot)$ 函数表示 sigmoid 函数, \mathbf{r}_t 表示重置门函数, \mathbf{z}_t 表示更新门函数, \mathbf{h}_t 表示当前时间步输出的滑坡特征信息,GRU 基本的单元结构图如图 4 所示。

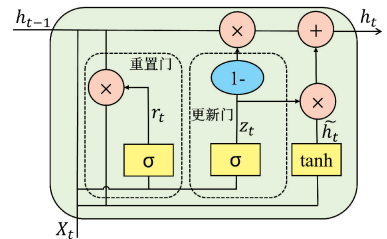


图 4 GRU 结构图

Fig. 4 Gated recurrent unit architecture diagram

2) GAM 优化 BiGRU

全局注意力机制关注输入数据的每一部分,通过对整个输入序列进行加权处理,帮助模型捕捉全局范围内的关键信息^[16]。GAM 优化 BiGRU 模型使其能够聚焦于输入序

列中最相关的部分,准确的理解数据的整体结构和内容,从而识别出重要的特征和模式,GAM 优化过程如下:

首先,将 BiGRU 的 s 时刻输出的隐藏状态记为 \mathbf{h}_s ,最终获得的目标隐藏状态记为 $\overline{\mathbf{h}}_s$ 。使用打分函数获得每时刻的得分,如式(10)所示。

$$\text{score}(\mathbf{h}_s, \overline{\mathbf{h}}_s) = \mathbf{h}_s^T \overline{\mathbf{h}}_s \quad (10)$$

通过 softmax 函数,将所有时刻得分进行归一化,从而得到各时刻权重 $a_t(s)$,如式(11)所示。

$$a_t(s) = \frac{\exp(\text{score}(\mathbf{h}_s, \overline{\mathbf{h}}_s))}{\sum_{s=1}^{T_x} \exp(\text{score}(\mathbf{h}_s, \overline{\mathbf{h}}_s))} \quad (11)$$

将各时刻的权重 $a_t(s)$ 与状态 \mathbf{h}_s 数乘计算,就得到了添加了 Global Attention 之后的上下文向量,最后每个时刻的求和后得到整个隐藏层状态的加权表示 \mathbf{c}_t 。每个时间步中,GAM 计算一个权重向量,表示模型对输入序列各部分的关注程度,权重向量应用于 BiGRU 输出的特征表示,通过对所有位置的特征进行加权,搜索调整其全局的注意力程度,对重要的部分赋予较大的注意力权重,对于影响小的部分赋予较小的注意力权重,就可以实现关注滑坡数据中重要的时域特征,提高 BiGRU 网络的对滑坡特征数据的感知能力,如式(12)所示。

$$\begin{cases} \mathbf{c}_t = \sum_{s=1}^{T_x} a_t \cdot \overline{\mathbf{h}}_s \\ \tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_c [\mathbf{c}_t; \mathbf{h}_t]) \end{cases} \quad (12)$$

式中: T_x 为滑坡数据时间序列长度, \mathbf{W}_c 为权值矩阵, $\tilde{\mathbf{h}}_t$ 为全局注意力机制的输出隐藏状态。全局注意力机制的结构示意图如图 5 所示。

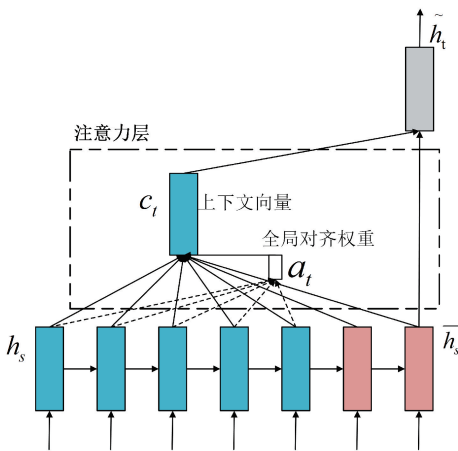


图 5 全局注意力机制结构图

Fig. 5 Structure diagram of the global attention mechanism

1.5 CAM 信息融合

交叉注意力机制 CAM 主要用于不同输入之间的信息交互^[17],有效地将来自不同来源的上下文进行对齐和关注,从而帮助模型更好地捕捉两个输入之间的相关性。

将上述过程 Transformer 编码器分支和 BiGRU-GAM 分支并行处理得到的滑坡全局与趋势特征信息进行特征融合,通过 CAM 融合输出。CAM 机制的核心在于计算两个不同输入序列之间的注意力权重,并根据这些权重来融合信息。具体的流程如下:

首先,基于查询 \mathbf{Q} 、键 \mathbf{K} 和值 \mathbf{V} 计算查询和键之间的相似度:在融合过程中将 Transformer 编码器分支的时序特征作为查询向量 \mathbf{Q} ,BiGRU-GAM 分支的提取的特征作为键值对序列 \mathbf{K} 和值 \mathbf{V} ,得到两个输入分支之间的相关性得分。这个得分反映了查询 \mathbf{Q} 与键 \mathbf{K} 在不同位置的相似程度。之后,通过 softmax 函数将这些相似度转换为概率分布,表示查询 \mathbf{Q} 对每个键 \mathbf{K} 的注意力权重。最后,将这些注意力权重应用到值 \mathbf{V} 上,最终得到输出向量。这相当于将关注的信息从值序列中提取出来,输入到下一个网络层。总体如式(13)所示。

$$\mathbf{y} = \text{softmax}((\mathbf{W}_Q \mathbf{F}_1)(\mathbf{W}_K \mathbf{F}_2)^T) \mathbf{W}_V \mathbf{F}_2 \quad (13)$$

式中: $\mathbf{F}_1, \mathbf{F}_2$ 分别为 Transformer 编码器和 BiGRU-GAM 处理后得到的滑坡特征信息, $\text{softmax}(\cdot)$ 表示归一化函数, \mathbf{W} 表示各项参数矩阵, \mathbf{y} 表示输出的滑坡位移预测值。

1.6 多级注意力机制并行模型

多级注意力机制并行模型的结构主要采用两个分支并行处理数据后再进行特征信息融合的方式。首先,原始数据经过数据预处理提高数据的可靠性以及加快模型的训练速度。之后,在 Transformer 编码器分支部分,主要使用 Transformer 编码器结构捕捉全局的时序模式,适合处理滑坡位移数据等长时间序列数据;在 BiGRU-GAM 分支部分,滑坡历史数据经 BiGRU 的处理后通过 GAM 增强的对关键数据的感知能力。最后经过两个并行分支处理后的滑坡特征数据,通过 CAM 进行特征融合并整合信息输出。这种多级注意力机制并行模型 (Transformer-BiGRU-GAM-CAM) 能够同时捕捉全局和依赖关系等特征信息,在滑坡位移多步预测方面有着显著的优势。总体的模型结构图如图 6 所示。由图 6 可知,该模型使用多头注意力机制、全局注意力机制并行进行信息提取与捕捉,使用交叉注意力机制进行特征信息的融合处理。多级注意力机制的模型 (Transformer-BiGRU-GAM-CAM) 能够聚焦信息的处理,对重点的影响因素和数据赋予更大的权重和关注,在数据处理和预测方面具有较大的优势。

2 实例分析

2.1 研究区与监测数据概况

广东省阳春市 K60+520~K60+800 段公路边坡地处高速公路附近,地形坡度较大,边坡以土质为主,易受到外界环境等因素的干扰,近年来该边坡发生长时间持续缓慢的位移变化。滑坡多步预测需要有长时间持续的边坡位移变化,捕捉长时间位移变化之间的依赖关系,从而提前预测

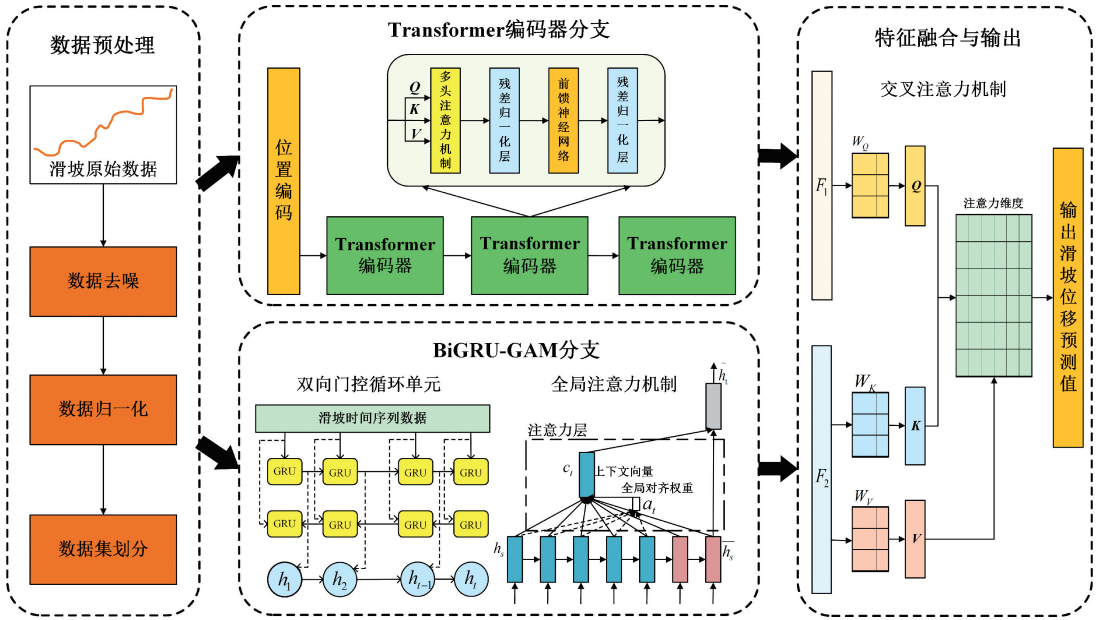


图 6 多级注意力机制并行模型结构图

Fig. 6 Architecture of the multi-level attention parallel model

滑坡可能进入临滑点的时间,尽早做出预警和反应。该边坡两年内位移发生长时间变化,监测数据资料持续完整,可以较好的表现土质滑坡长时间的持续位移演变过程,反映了众多同类型土质滑坡发展的特点,适合于实验模型建模。因此,本文采用广东省阳春市 K60+520~K60+800 段公路边坡为例进行实验验证,采用监测设备采集该边坡的位移变化数据、降雨量、土壤含水率等数据进行分析。现场及监测站点图如图 7 所示。



图 7 公路边坡现场与监测站点图

Fig. 7 On-site diagram of highway slope and monitoring stations

GNSS 监测设备监测到的滑坡位移数据分为 X 方向(东西)、Y 方向(南北)和 Z 方向(高程)3 个方向的监测数据,监测数据的间隔为每小时监测一次,该边坡 2022 年 12 月 13 日~2024 年 8 月 20 日的 3 个方向的滑坡位移曲线图分别如图 8(a)~(c)所示。从图中可以看出该边坡的 X 方向和 Z 方向在 2024 年 3 月均发生了较大的连续的位移移动变化,且位移移动在长时间内持续发生。而 Y 方向则在较短的时间点处发生“脉冲”式的位移移动,但总体趋势

相对平稳。

采集到的滑坡原始数据往往含有多种高频噪声,为了尽量消除外界噪声活动的影响,通过滑动移动平均算法对滑坡原始数据进行去噪处理,剔除易于干扰位移的高频噪声,保留具有有效信息特征的低频位移项,数据去噪后的各方向位移结果图 8(a)~(c)所示。

2.2 滑坡监测数据处理

1) 数据预处理

要进行准确的多步位移预测,首先需要对原始数据进行影响因素的筛选,以确定合适的输入项,提高模型多步预测的稳定性。以预测 X 方向的位移数据为例,需要筛选出与 X 方向位移关联度高的监测项。经过筛选后的影响因素作为输入项存在数据量纲不一致的情况,影响模型预测的精度,还要通过数据归一化的处理以消除量纲的影响,提升模型预测数据的稳定性。

使用最小-最大归一化法对影响因素数据进行归一化处理,去除量纲对数据的影响。最小-最大归一化法如式(14)所示。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (14)$$

式中: x 为滑坡某项数据值, x_{\min} 为最小值, x_{\max} 为最大值, x' 为归一化后的滑坡数据值。

最后,将数据按照 7:1:2 的比例划分训练集、测试集以及验证集,用于模型训练、参数调整以及模型效果的验证。实验训练的 软件环境: 框架 pytorch2.2.1, 语言 python3.12, 计算架构为 CUDA12.1。训练的硬件平台: CPU 为 12400F, GPU 为 NVIDIA GeForce RTX4060Ti

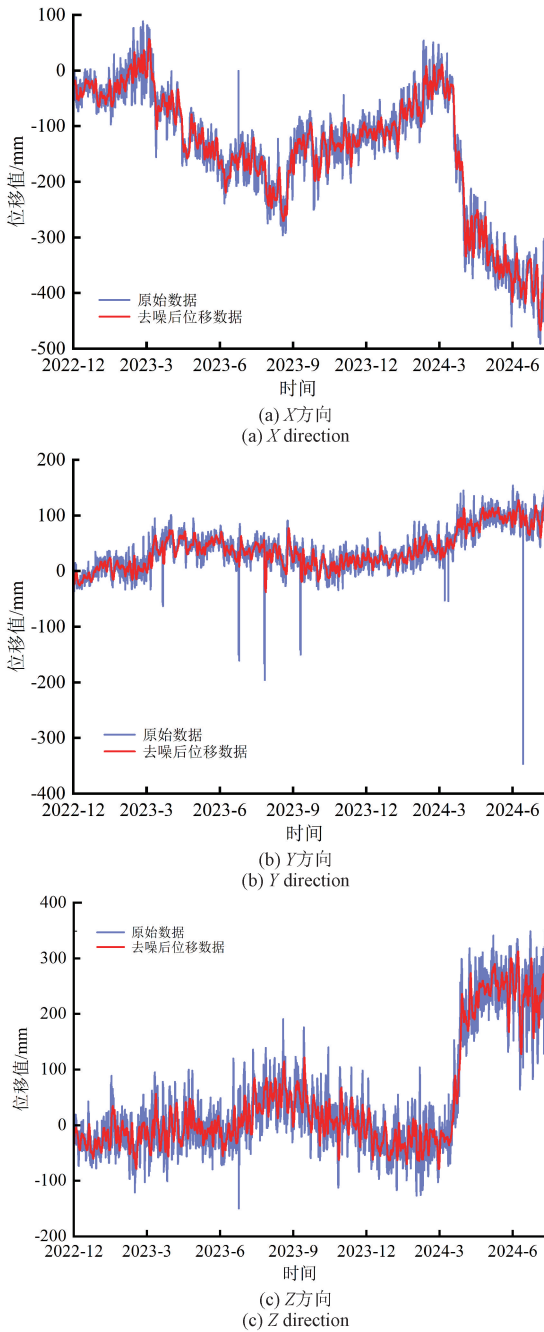


图 8 滑坡位移曲线图

Fig. 8 Landslide displacement curve graph

8 G 显卡。

2) 影响因素筛选

在滑坡位移预测中选择合理的影响因素可以提升预测模型的可靠性。灰色关联度分析 (grey relational analysis, GRA) 方法用灰色关联度值的大小来描述因素间关系的强弱, 保留关联性较强的影响因素^[18]。

以 X 方向位移预测为例, 选用该滑坡中的各监测项对其进行关联性分析, 得出各监测项对 X 方向位移值的关联度值如图 9 所示。

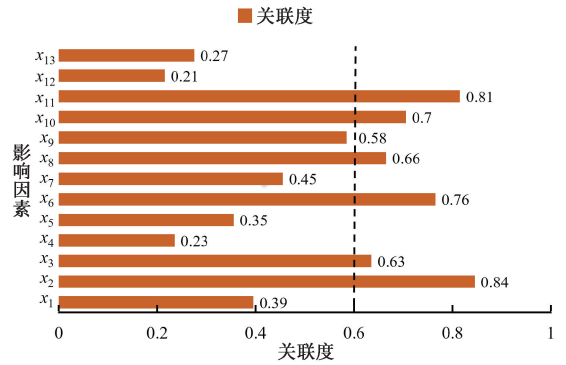


图 9 灰色关联度分析结果

Fig. 9 Results of grey relational analysis

$x_1 \sim x_{13}$ 分别表示 Y 方向位移、Z 方向位移、倾角 (X、Y、Z 方向)、加速度 (X、Y、Z 方向)、裂缝位移、土壤含水率、降雨量、超声、次声。通过分析这 13 项监测值与 X 方向位移的关联度, 选取合适因素作为输入项。

当影响因素与滑坡 X 方向位移的关联度超过 0.6 时, 认为该影响因素与滑坡变形有着较强的相关性。由图 8 可知, Z 方向位移、X 方向倾角、X 方向加速度、Z 方向加速度、土壤含水率、降雨量与 X 方向位移有着较强相关性。因此, 将上述几项影响因素作为滑坡 X 方向位移预测的输入因素。

2.3 评价指标

为了评价模型的性能好坏, 本文采用平均绝对误差 (mean absolute error, MAE)、均方误差 (mean square error, MSE)、均方根误差 (root mean square error, RMSE)、决定系数 (coefficient of determination, R^2) 对预测数据进行量化分析, 作为模型的评价指标。MAE、MSE、RMSE 的值越低, R^2 的值越高, 表明模型的性能越好, 预测的准确度越高, 表现更稳定。各评价指标的计算如式 (15)~(18) 所示。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (16)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (17)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (18)$$

式中: y_i 为第 i 个真实值, n 为样本数, \hat{y}_i 为位移预测值, \bar{y} 为位移平均值。

2.4 结果分析

1) 模型超参数寻优

调整模型的超参数, 可以提高模型性能和预测精度。

本文通过调整 Transformer 编码器层数(范围[1,100])、多头注意力层数(范围[1,100])、BiGRU 隐藏层神经元个数(范围[100,1 000])、学习率(范围[0.000 01,0.1])、最大训练次数(范围[50,1 000])来进行寻优。这些超参数的优劣影响了模型规模是否复杂、模型的训练效果是否显著、模型的收敛速度是否迅速,最优超参数可以显著提升模型的性能。本文通过采用网格搜索法(grid search)进行超参数寻优,寻找最优的超参数组合,经过优化后的最佳参数结果如表 1 所示。

2) 滑坡位移预测

为了对比验证模型结构的有效性,使用在滑坡位移预测领域应用广泛的 BiGRU 模型、Transformer 模型、Transformer-BiGRU 模型和本文所提出的多级注意力机

表 1 模型最优参数值

Table 1 Optimal parameter values of the model

Transformer 编码器层数	多头注意力层数	隐藏层神经元个数	学习率	最大训练次数
6	8	120	0.000 3	100

制并行模型(Transformer-BiGRU-GAM-CAM)进行对比。通过对比验证 GAM 优化和 CAM 特征融合在滑坡位移数据多步预测中的有效性。

在滑坡位移多步预测实验中,分别进行预测时间步长为 3、5、7、9 步以及更长步 20 步的情况。以 X 方向为例,各模型一次预测 3、5、7、9 步时间步长的预测结果对比如图 10 所示。

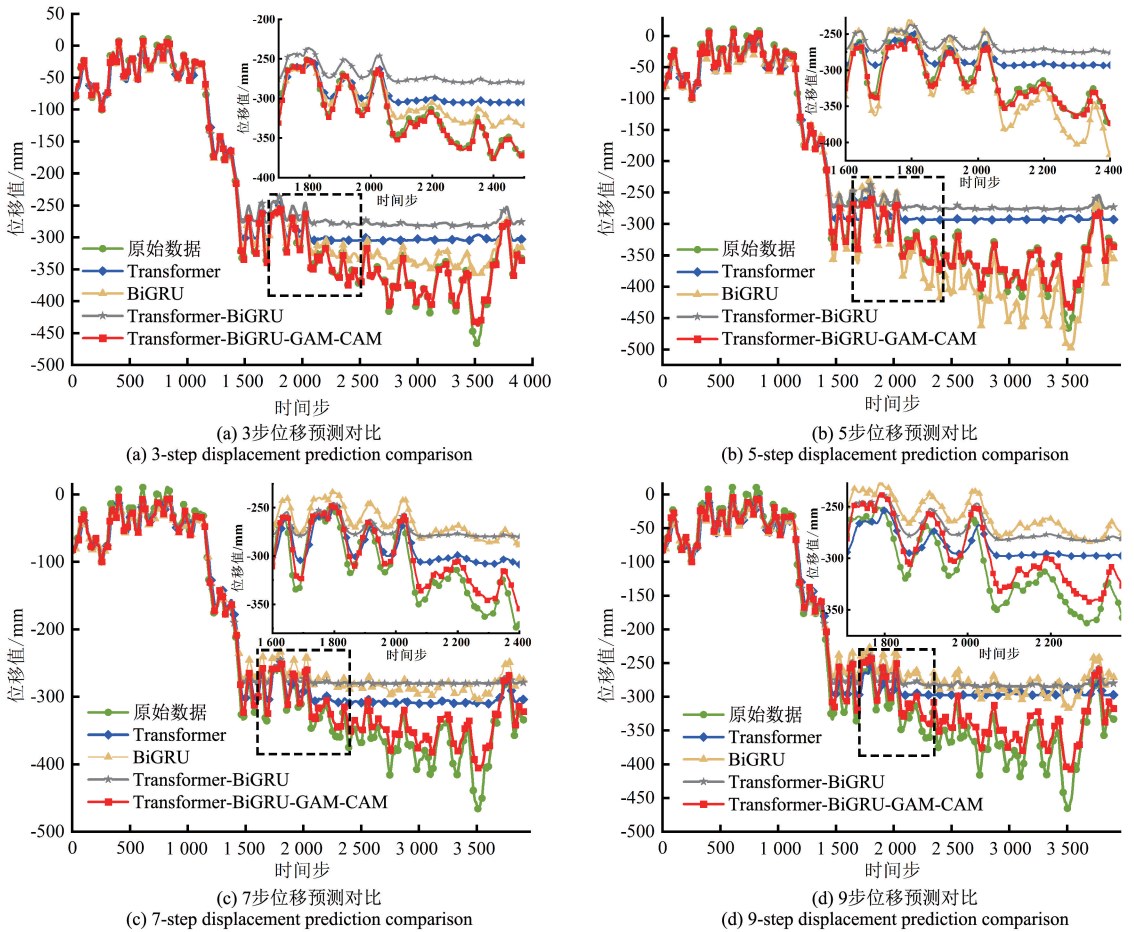


图 10 X 方向滑坡位移预测结果对比图

Fig. 10 Comparison chart of predicted results for landslide displacement in the X direction

在预测过程中,多步预测一次性预测输出多个时间步的值,输出 3、5、7、9 步的预测值其各自输入的时间序列的长度也是不同的,以下分别采用的是 12、20、28、36 的输入时间序列长度。从图 10 中可以看出,在预测的起始阶段,各模型的预测准确度相差不大,均能保持较好的拟合度。但随着时间的推移,在 3、5、7、9 各步上,BiGRU 模型、

Transformer 模型、Transformer-BiGRU 模型会出现较大的误差,模型的稳定性和预测准确率难以保证,而 Transformer-BiGRU-GAM-CAM 模型仍能够保持较高的准确度,虽然其随着预测步长的增加,准确度也出现了下降的情况,但是与其他模型的难以预测趋势的情况相比,Transformer-BiGRU-GAM-CAM 模型仍能够预测出变化

趋势,同时从图中也能得知,单纯采用 Transformer-GAM 并行模型进行多步预测,其结果相较 Transformer、BiGRU 模型反而会出现预测精度下降,而通过多级注意力机制 GAM 和 CAM 的优化融合后,多步预测精度大幅度提高,进一步验证了 GAM 优化和 CAM 特征融合后的模型在多

步滑坡位移数据预测方面的精度提升。

为了更直观的比较和评价各模型,使用平均绝对误差(MAE)、均方误差(MSE)、均方根误差(RMSE)、决定系数(R^2)来对模型进行评价。以 9 步预测时间步长为例,各模型评价指标如表 2 所示。

表 2 各模型预测评价指标对比

Table 2 Comparison of prediction evaluation metrics across different models

模型	MAE	MSE	RMSE	R^2
Transformer	5.937 8	41.589 6	6.449 0	0.868 9
BiGRU	6.646 9	56.943 6	7.546 1	0.835 0
Transformer-BiGRU	6.774 3	73.774 4	8.589 2	0.799 4
Transformer-BiGRU-GAM-CAM	2.168 2	9.319 0	3.052 7	0.968 9

从表 2 中可以看出,Transformer-BiGRU-GAM-CAM 模型的 MAE、MSE、RMSE 均是最低的,即误差最小; R^2 为 0.968 9,即模型拟合度最高。

在 Transformer-BiGRU-GAM-CAM 模型中,采用了多级的注意力机制作用,起到了优化 BiGRU 和融合特征的作用。特别是 CAM 能够在多特征中提取出重要的特征,增强特征的表示能力来实现高精度的预测。由于模型采用多输入多输出的预测策略,通过交叉注意力机制(CAM)能从序列特征中提取出对模型预测重要特征。为了分析各输入项在模型训练过程中的特征贡献度,将相关度高的输入特征进行特征可视化,特征贡献度结果如图 11 所示。

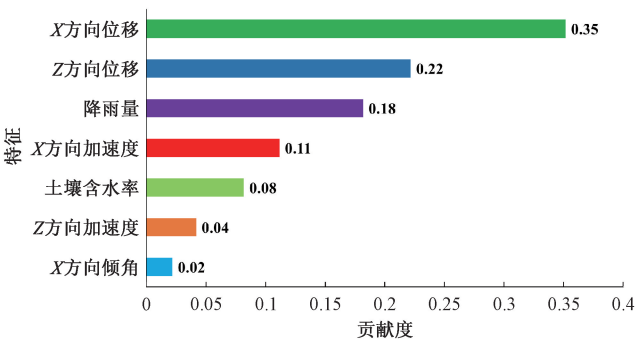


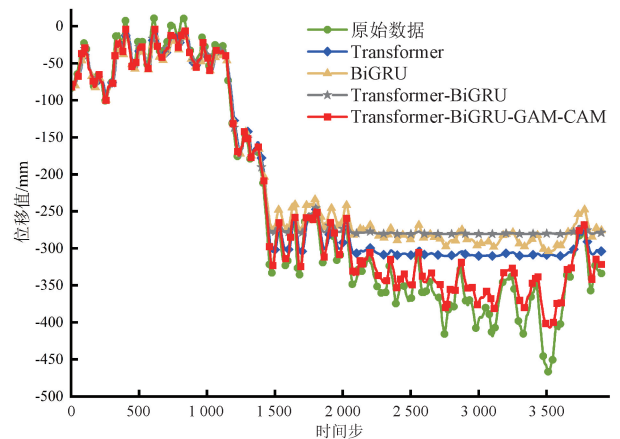
图 11 特征贡献度分析可视化

Fig. 11 Visualization of feature contribution analysis

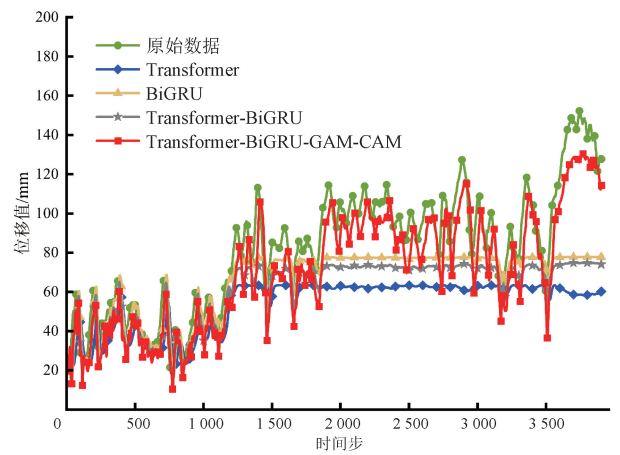
由图 11 可知,对 X 方向位移预测影响最大的是其本身的历史数据,Z 方向位移和降雨量数据也是影响预测位移的重要因素。特征贡献度与之前的灰度关联度分析结果吻合,验证了 CAM 能够在滑坡位移多步预测中提取出重要的影响特征,提升模型的性能。

为了验证在更长时间步长下,各模型预测效果的稳定性,使用各模型以 20 步的预测步长对 X、Y、Z 三个方向的数据进行预测,各模型预测结果对比如图 12 所示。

由 20 步的预测结果比较可以看出,随着预测时间的增加,特别是滑坡位移值出现较大的波动时,Transformer 模型、GRU 模型、Transformer-BiGRU 模型等均出现了难以拟合的情况,但本文所以提出的 Transformer-BiGRU-GAM-CAM 模型仍能够保持一定的拟合度,模型的预测准确度和稳定性效果较好。



(a) X 方向
(a) X direction



(b) Y 方向
(b) Y direction

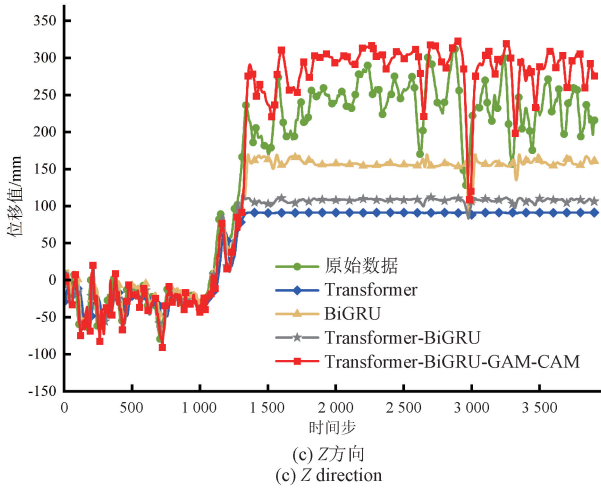


图 12 滑坡位移 20 步预测结果对比图

Fig. 12 Comparison chart of 20-step forecast results for landslide displacement

3 结 论

本文以滑坡位移监测工作为背景,结合滑坡位移多步预测应用场景及需求,使用多级注意力机制相结合的结构,提出一种基于多级注意力机制并行模型(Transformer-BiGRU-GAM-CAM)的滑坡多步位移预测方法,并在广东省阳春市 K60+520~K60+800 公路边坡为实例进行了预测验证,实验效果良好,说明了模型的可行性与有效性。

在滑坡位移多步预测中,Transformer-BiGRU-GAM-CAM 模型与 GRU 模型、Transformer 模型、Transformer-BiGRU 模型比较中 MAE、RMSE 结果均最低,9 步预测结果分别为 2.17 mm、3.05 mm;在决定系数(R^2)的比较中,Transformer-BiGRU-GAM-CAM 模型为 0.968 9,更接近于 1,准确率更高,模型表现更稳定,同时使用特征贡献度分析证明了 GAM 优化和 CAM 特征融合在优化滑坡位移多步预测中的有效性。

采用多输入多输出的预测策略,在滑坡位移多步预测中相较于直接预测和滚动预测表现出更好的多步预测性能,能够降低直接预测中的模型复杂度和避免滚动预测中累积误差的问题,在长时间步预测的表现中更加稳定可靠。

Transformer-BiGRU-GAM-CAM 模型在更长预测时间步下具有较好的稳定性,能够更准确及时的进行滑坡位移预测,利用多输入多输出的预测策略,在 3、5、7、9 步预测均具有不错的效果,在更长的 20 步预测中保持的效果也更稳定,提升了滑坡位移多步预测的准确度,有利于提前预知滑坡发展动向,为滑坡的预防与治理提供了重要的技术支持。

参 考 文 献

[1] 唐王菲,周海莲,唐天俊,等. 融合动静态变量的滑坡

多步位移预测方法[J]. 吉林大学学报(工学版), 2023(6):1833-1841.

TANG F F, ZHOU H L, TANG T J, et al. Multi-step prediction method of landslide displacement based on fusion dynamic and static variables[J]. Journal of Jilin University(Engineering and Technology Edition), 2023(6):1833-1841.

[2] 刘勇,秦志萌,余宏明,等. 基于 ENN 广义预测控制算法的滑坡位移多步预测[J]. 地质科技通报, 2017, 36(4):255-259.

LIU Y, QIN ZH M, YU H M, et al. Generalized predictive control algorithm for landslide displacement multi-step prediction [J]. Geological Science and Technology Information, 2017,36(4):255-259.

[3] 宋星辰. 基于深度学习的时间序列多步预测的研究与应用[D]. 荆州:长江大学,2022.

SONG X CH. Research and application of time series multi-step forecasting based on deep learning [D]. Jingzhou:Yangtze University, 2022.

[4] 雷德鑫,易武. 基于 ARIMA 时间序列模型的滑坡位移预测预报——以三峡库区王家坡滑坡为例[J]. 人民长江, 2018,49(21):56-60,83.

LEI D X, YI W. Prediction and forecast of landslide displacement based on ARIMA model; Case of Wangjiapo landslide in Three Gorges Reservoir area[J]. Yangtze River, 2018,49(21):56-60,83.

[5] 汪标,易庆林,邓茂林,等. 基于累积位移特征与时间序列组合模型的滑坡位移预测[J]. 工程地质学报, 2024, 32(5):1629-1639.

WANG B, YI Q L, DENG M L, et al. Landslide displacement prediction based on cumulative displacement features and time series combination model[J]. Journal of Engineering Geology, 2024, 32(5):1629-1639.

[6] MENG SH Q, SHI ZH M, PENG M, et al. Landslide displacement prediction with step-like curve based on convolutional neural network coupled with bi-directional gated recurrent unit optimized by attention mechanism[J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108078.

[7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30, DOI: 10.5555/3295222.3295349.

[8] 唐海林,张俊,李屹旭,等. 基于注意力改进 RTformer 的滑坡遥感图像语义分割[J]. 电子测量技术, 2024, 47(21):150-158.

TANG H L, ZHANG J, LI Y X, et al. Semantic

- segmentation of landslide remote sensing image based on improved attention RTformer [J]. *Electronic Measurement Technology*, 2024, 47(21):150-158.
- [9] 罗祎沅, 蒋亚楠, 许强, 等. 基于最优分解模态和 GRU 模型的库岸滑坡位移预测研究[J]. *武汉大学学报(信息科学版)*, 2023, 48(5):702-709.
- LUO H Y, JIANG Y N, XU Q, et al. Displacement prediction of reservoir bank landslide based on optimal decomposition mode and GRU model[J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(5):702-709.
- [10] GE Q, LI J, WANG X H, et al. LiteTransNet: An interpretable approach for landslide displacement prediction using transformer model with attention mechanism [J]. *Engineering Geology*, 2024, 331:107446.
- [11] 田原, 庞骁, 赵文祎, 等. 基于 Transformer 的滑坡短期位移预测模型[J]. *北京大学学报(自然科学版)*, 2023(2):197-210.
- TIAN Y, PANG X, ZHAO W Y, et al. A transformer-based model for short-term landslide displacement prediction [J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2023 (2): 197-210.
- [12] 李斌, 舒嘉辉, 严灵潇, 等. 改进黑翅鸢算法的 1D-2D-GAF-PCNN-GRU-MSA 弓网电弧检测应用[J]. *电子测量与仪器学报*, 2024, 38(10):201-211.
- LI B, SHU J H, YAN L X, et al. 1D-2D-GAF-PCNN-GRU-MSA pantograph arc detection application based on improved black-winged kite algorithm[J]. *Journal of Electronic Measurement and Instrumentation*, 2024, 38(10):201-211.
- [13] NAVA L, CARRARO E, REYES-CARMONA C, et al. Landslide displacement forecasting using deep learning and monitoring data across selected sites[J]. *Landslides*, 2023, 20(10): 2111-2129.
- [14] CHEN Z L, MA M B, LI T R, et al. Long sequence time-series forecasting with deep learning: A survey [J]. *Information Fusion*, 2023, 97:101819.
- [15] 李欣宇, 孙传猛, 魏宇, 等. 融合 Transformer 与残差通道注意力的恶劣场景水位智能检测方法[J]. *电子测量与仪器学报*, 2023, 37(1):59-69.
- LI X Y, SUN CH M, WEI Y, et al. Water level intelligent detection method based on fuse Transformer residual channel attention mechanism in harsh environments[J]. *Journal of Electronic Measurement and Instrumentation*, 2023, 37(1):59-69.
- [16] LI Y Y, ZHOU ZH G, QI G Q, et al. Remote sensing micro-object detection under global and local attention mechanism [J]. *Remote Sensing*, 2024, 16(4):644.
- [17] LI H, WU X J. CrossFuse: A novel cross attention mechanism based infrared and visible image fusion approach[J]. *Information Fusion*, 2024, 103:10214.
- [18] 封青青, 李丽敏, 陈飞阳, 等. 基于 ICEEMDAN 分解与 SE 重构和 DBO-LSTM 的滑坡位移预测[J]. *电子测量技术*, 2024, 47(7):80-87.
- FENG Q Q, LI L M, CHEN F Y, et al. Landslide displacement prediction based on ICEEMDAN decomposition and SE reconstruction and DBO-LSTM[J]. *Electronic Measurement Technology*, 2024, 47 (7): 80-87.

作者简介

任冯, 硕士研究生, 主要研究方向为滑坡监测预测技术。

E-mail: 487641079@qq.com

肖慧(通信作者), 硕士, 副教授, 主要研究方向为光栅光纤在地质灾害监测领域的应用研究。

E-mail: 200860050@ecut.edu.cn