

DOI: 10.16285/j.rsm.2021.2091

基于多头自注意力机制的 LSTM-MH-SA 滑坡位移预测模型研究

张振坤¹, 张冬梅¹, 李 江², 吴益平³

(1. 中国地质大学 计算机学院, 湖北 武汉 430074; 2. 湖北省自然资源厅信息中心, 湖北 武汉 430071;
3. 中国地质大学 工程学院, 湖北 武汉 430074)

摘 要: 受自身地质条件及外界周期、随机等因素影响, 滑坡演变过程具有典型跃变特征。传统基于门控机制的深度学习模型对阶跃型滑坡预测能力不足, 多头自注意力通过关注不同尺度时序数据的隐含信息能自适应挖掘序列的变化程度特征, 有效学习数据潜在变化趋势, 提升序列的预测能力。研究基于变分模态分解技术将滑坡累积位移量分解成趋势项、周期项和随机项, 对各位移分量和影响因子开展动态时间规整相关性分析。结合多头自注意力机制和长短时记忆网络模型对各位移分量进行动态预测, 各位移分量预测值相加得到实际预测结果。以三峡库区白水河滑坡作为研究区, 对监测点 ZG118 开展累积位移预测, 采用监测点 ZG93、XD01 进行模型适应性验证, 试验结果表明对于降雨、库水位变化导致的阶跃数据段, 新模型能大大提升预测的精度, 为三峡库区滑坡位移预测研究提供新的思路。

关 键 词: 滑坡位移预测; 变分模态分解; 动态时间规整; 多头自注意力机制; 长短时记忆网络

中图分类号: P642

文献标识码: A

文章编号: 1000-7598(2022) 增刊 2-0477-11

LSTM-MH-SA landslide displacement prediction model based on multi-head self-attention mechanism

ZHANG Zhen-kun¹, ZHANG Dong-mei¹, LI Jiang², WU Yi-ping³

(1. School of Computer Science, China University of Geosciences, Wuhan, Hubei 430074, China; 2. Information Center, Department of Natural Resources of Hubei Province, Wuhan, Hubei 430071, China; 3. Faculty of Engineering, China University of Geosciences, Wuhan, Hubei 430074, China)

Abstract: Under the influences of the geological conditions of landslide and external periodic and random factors, the evolution process of the landslide has typical abrupt characteristics. The conventional deep learning model based on gating mechanism has insufficient ability to predict step mutation sequence, while by mining the potential information of time series data with different scales, multi-head self-attention can adaptively find the alteration trend of sequences and improve the prediction ability of landslide sequence. Based on variational mode decomposition, the cumulative displacement of landslide was decomposed into trend displacement, periodic displacement and stochastic displacement. Dynamic time warping algorithm was used to analyze the correlation between each component and influencing factors. These displacement components were dynamically predicted by a modified model integrated with long short-term memory (LSTM) neural network and multi-head self-attention mechanism, and the predicted values of each component were added to obtain the cumulative displacement prediction result. The monitoring point ZG118 of Baishuihe landslide in the Three Gorges Reservoir area was taken as an example to predict the cumulative displacement in this paper. The monitoring points ZG93 and XD01 were used for model adaptability verification. The results show that the new model can greatly improve the prediction accuracy for the step data mutation caused by the changes of rainfall and reservoir water level, providing a new idea for the prediction of landslide displacement in the Three Gorges Reservoir area.

Keywords: landslide displacement prediction; variational mode decomposition; dynamic time warping; multi-head self-attention mechanism; long short-term memory (LSTM) neural network

收稿日期: 2021-12-10

修回日期: 2022-07-21

基金项目: 国家自然科学基金联合基金重点支持项目 (No.U1911205); 湖北省自然资源厅科技项目 (No.ZRZY2022KJ07); 湖北省自然科学基金 (No.2019CFA023)。

This work was supported by the Key Projects Supported by NSFC Joint Fund (U1911205), the Science and Technology Project of Department of Natural Resources of Hubei Province (ZRZY2022KJ07) and Hubei Natural Science Foundation of China(2019CFA023).

第一作者简介: 张振坤, 男, 1998 年生, 硕士研究生, 主要从事滑坡智能预测方面的研究。E-mail: 1202111251@cug.edu.cn

通讯作者: 张冬梅, 女, 1972 年生, 博士, 教授, 博士生导师, 主要从事地学数据挖掘、滑坡智能预测的教学和科研工作。E-mail: cugzdm@foxmail.com

1 引言

滑坡是一种发生较为频繁且危害性较大的地质灾害,对人类生命财产和环境造成巨大破坏。预测预报是预防滑坡灾害的有效手段,滑坡位移预测特别针对三峡库区广泛分布的阶跃型滑坡是研究的热点和难点^[1]。传统研究主要是针对数据的非线性非稳态特征,将滑坡累积位移量进行时频分解,对位移分量分别进行预测^[2]。

分解非平稳滑坡位移量的方法主要有两类,一类结合数学方法如移动平均法和平移指数法等分离出物理意义明确的趋势项和周期项^[3],但无法分解出随机项;另一类基于小波分析、模态分解技术等将滑坡位移量分解为若干具有不同特征的位移分量^[4],小波分析往往难以确定小波基,自适应分解方法虽在一定程度上能提高预测精度,但各分量的物理意义不够明确。研究采用变分模态分解技术 (variational mode decomposition, 简称 VMD) 对滑坡位移量开展自适应分解,以对不同尺度时序分量和地质、降雨、库水位等多影响因素进行相关性分析。

预测模型包括静态模型和动态模型。静态模型如张俊^[5]、Zhou^[6]等采用结合粒子群优化和支持向量机回归的 PSO-SVR 模型对滑坡周期项进行预测,李麟玮等^[7]提出改进支持向量机 GWO-MIC-SVR 模型,檀梦皎等^[8]提出结合粒子群优化和极限学习机的 PSO-ELM 模型等。上述静态模型忽视了滑坡演化过程是个复杂的动态系统,难以有效提取历史序列的长程相关特征。动态模型考虑了数据的时效性和序列变化特征,如杨背背等^[9]构建的长短期记忆网络 (long and short term memory neural network, 简称 LSTM) 模型对周期项位移进行预测,罗祎沅等^[10]提出基于门控循环单元 (gate recurrent unit, 简称 GRU) 的滑坡位移分量预测。这些模型虽然实现了网络状态的动态反馈,但基于循环神经网络引入的门控机制和记忆单元改进网络模型,未能从根本上解决门控组件难以学习跃变数据变化特征的问题,对局部变化剧烈的阶跃位移段数据预测效果较差。2017 年 Vaswani 等^[11]提出多头自注意力机制 (multi-head self-attention mechanism, 简称 MH-SA), 区分不同尺度信息特征并学习融合,能刻画各种内外诱发因素导致的局部滑坡变化特征,提升模型预测性能。

论文以三峡库区白水河滑坡为研究区,采用变

分模态分解将累积位移曲线分解为具有不同物理意义及尺度特征的分量,结合多头自注意力机制和长短时记忆网络模型设计 LSTM-MH-SA 新模型进行预测,各分项预测值相加得到最终滑坡累积位移预测值。新模型通过多头自注意力对不同子空间的特征进行提取,对重要特征加权关注和学习,捕捉整个序列内部的依赖关系,强化输出影响较大参数的权重,能较好提取不同尺度数据的特征,进一步提升滑坡位移预测的效果。

2 滑坡位移预测模型

滑坡累积位移数据是典型非线性时间序列。针对不同尺度数据特征,滑坡累积位移定义如下:

$$Z(t) = p(t) + q(t) + r(t) \quad (1)$$

式中: $Z(t)$ 为 t 时刻的滑坡累积位移; $p(t)$ 为 t 时刻的趋势项位移,反映长期变化趋势,受滑坡自身地质条件 (如地质构造、岩土体组成等) 影响总体趋势呈单调递增变化; $q(t)$ 为 t 时刻的周期项位移,反映外界诱发因素如库水位人工调控导致周期性变化特征; $r(t)$ 为 t 时刻的随机项位移,刻画非季节性降雨等引起的滑坡位移量随机性变化。由于监测手段不足,随机噪声往往忽略不计。

2.1 变分模态分解

变分模态分解是 2014 年 Dragomiretskiy 等^[12]在经验模态分解 (empirical mode decomposition, 简称 EMD) 的基础上提出的一种自适应的、完全非递归的模态变分和信号处理方法,能较好克服 EMD 存在的端点效应和模态混叠问题。

VMD 分解实质是约束变分问题的求解过程,将原始信号 f 分解为 K 个固有模态分量 (intrinsic mode function, 简称 IMF), 约束变分表达式为

$$\min_{\{u_k, w_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\sigma(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j w_k t} \right\|_2^2 \right\} \quad (2)$$

$$\sum_{k=1}^K u_k = f(t) \quad (3)$$

式中: $[\sigma(t) + j / \pi t] u_k(t)$ 为各分量的解析信号; $e^{-j w_k t}$ 为各信号的预估中心频率; $f(t)$ 为原始信号。

引入二次项惩罚系数与 Lagrange 乘子 λ , 转换为非约束变分问题对模态分量 u_k 、中心频率 w_k 进行求解。VMD 求解算法见表 1。

通过 VMD 分解得到滑坡位移量序列不同层次的信息,并对应实际物理意义。

表 1 VMD 算法流程
Table 1 Algorithm flow of VMD

算法: VMD 求解算法
输入: 原始信号 f
输出: K 个 IMF 分量
Step1 初始化模态 u_k , 中心频率 w_k , 拉格朗日乘数 λ 、迭代次数 $n=0$ 、最大频数 K 。
Step2 开始迭代过程, $n=n+1$, $k=1$ 。
Step3 更新 u_k , w_k 和 λ , $k=k+1$ 。
Step4 判断 k 是否达到最大频数 K (k 为已更新的模态分量的数量), 是则转向 step5, 否则转向 step3。
Step5 设置精度 ε , 判断是否满足精度条件
$\frac{\sum_{k=1}^K \ u_k^{(n+1)} - u_k^n\ _2^2}{\ u_k^n\ _2^2} < \varepsilon$
若满足则停止迭代, 输出 K 个 IMF 分量, 若不满足转向 step2。

2.2 动态时间规整

动态时间规整 (dynamic time warping, 简称 DTW) [13] 是基于动态规划思想, 整合时间弯曲和距离计算的非线性序列相似性度量技术, 能克服时间轴偏移和拉伸以消除滞后性产生的偏差, 可量化库水位、降雨量等因素对滑坡位移量变化的影响。

已知两个序列 $X = \{x_1, x_2, \dots, x_{|X|}\}$ 与 $Y = \{y_1, y_2, \dots, y_{|Y|}\}$, 时间规整函数 C 定义为

$$C = \{c_1, c_2, \dots, c_n, \dots, c_{|X||Y|}\} \quad (4)$$

式中: c_n 为序列间的匹配点对 (x_i, y_j) , 匹配点间的距离为局部匹配距离。采用动态规划递归求解搜索路径, 满足路径上所有匹配点对间的局部匹配距离总和最小。

根据序列 $|X|$ 、 $|Y|$ 的长度, 生成由 $|X| \cdot |Y|$ 个基距离 $d(i, j)$ 构成的全局搜索矩阵, 在满足单调性、有界性、连续性和边界限制的约束条件下搜索计算累加距离 $D(i, j)$, 路径搜索策略如图 1 所示。

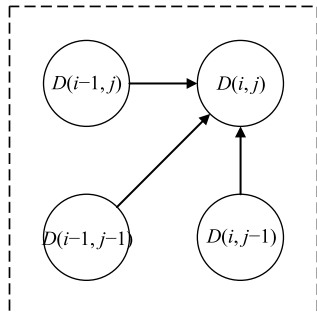


图 1 DTW 路径搜索示意图
Fig.1 Path search for DTW

累加距离 $D(i, j)$ 为 $D(i-1, j)$ 、 $D(i, j-1)$ 和 $D(i-1, j-1)$ 3 个方向最小累加距离与该点的基距离 $d(i, j)$ 之和, 其状态转移方程表示如下:

$$D(i, j) = d(i, j) + \min \begin{cases} D(i-1, j) \\ D(i, j-1) \\ D(i-1, j-1) \end{cases} \quad (5)$$

DTW 距离是以 $(1, 1)$ 为起点搜索到终点 $(|X|, |Y|)$ 的最小累加距离。考虑到匹配元素数量以及取值区间的影响, 以 DTW 归一化序列输入的平均距离作为衡量曲线相似度的度量, DTW 值越小相关性越高。

2.3 长短时记忆网络

长短时记忆网络是由 Hochreiter 等 [14] 在 1997 年为解决循环神经网络的长距离依赖问题提出的添加了门控机制和记忆单元的改进网络。LSTM 由记忆单元和输入门、输出门以及遗忘门等门控单元组成, 结构如图 2 所示。

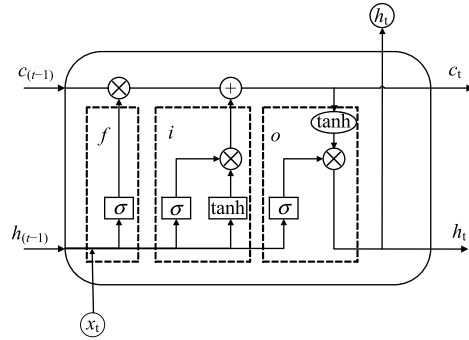


图 2 LSTM 结构图
Fig.2 Architecture of LSTM neural network

t 时刻隐藏层状态传递过程为

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (6)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (7)$$

$$a_t = \tanh(W_a[h_{t-1}, x_t] + b_a) \quad (8)$$

$$c_t = c_{t-1} \odot f_t + i_t \odot a_t \quad (9)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (10)$$

$$h_t = o_t \odot \tanh(c_t) \quad (11)$$

式中: f_t 、 i_t 、 o_t 分别为遗忘门、输入门与输出门在 t 时刻的向量值; W_f 、 W_i 、 W_o 、 W_a 与 b_f 、 b_i 、 b_o 、 b_a 为各门控单元和候选记忆状态的系数矩阵与偏置矩阵; a_t 为候选记忆状态; σ 为 sigmoid 激活函数, c_t 、 c_{t-1} 与 h_t 、 h_{t-1} 为当前时刻和前一时刻记忆单元状态和隐藏层状态; $\odot$ 表示元素对应相乘。通过门控机制更新记忆单元状态和隐藏层状态传递信息。

LSTM 预测模型在一定程度上解决了长距离依赖问题, 但由于记忆长度有限, 门控机制对于阶跃

型数据局部变化特征学习较弱, 难以有效预测。

2.4 多头自注意力机制

注意力机制由 1980 年 Treisman^[15]提出, 通过计算输入信息的注意力概率分布关注重点特征, 强化对输出影响较大的参数, 增强模型预测能力。自注意力机制是由谷歌团队提出的一种特殊注意力机制, 对序列自身通过计算注意力求取内部自相关权重, 关注重要特征, 融入样本局部动态信息, 提高预测精度, 如图 3 所示。

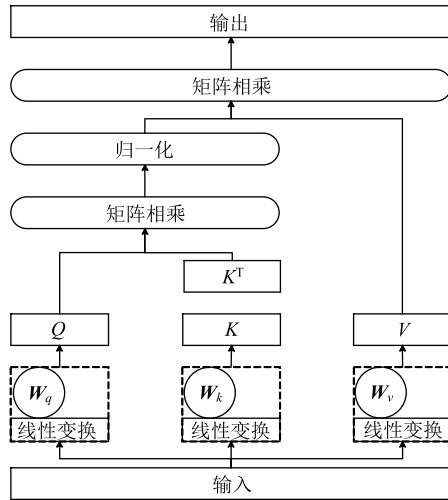


图 3 自注意力机制结构图

Fig.3 Structure of self-attention mechanism

W_q 、 W_k 、 W_v 是对输入进行线性映射的系数矩阵, 输入序列与系数矩阵相乘分别得到 Q (query)、 K (key)、 V (value), 与注意力机制不同, 自注意力中 $Q=K=V$, Q 为查询序列, $K-V$ 是键值对, 通过点积运算, 求 Q 和 K 的相似度。 d_k 为各键的特征维度, 用于权重缩放, 经 softmax 归一化至 [0, 1] 间, 得到 K 中各键权重并与真实值 V 相乘, 输出融合自相关权重的特征矩阵, 如下式所示:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

多头自注意力机制是对自注意力机制的一种改进, 将自注意力机制分成多个平行的头进行多次线性映射, 对不同映射结果进行 Attention 计算, 处理输入序列中不同层次的信息, 计算全局注意力, 模型结构如图 4 所示。其实质是多个独立的自注意力集成。将时间序列分为 h 个子空间, 各头对子空间进行自注意力计算, 增强注意力的表达能力, 拼接、集成 h 个头的结果提取特征。各头分别关注不同层次时序数据的隐含信息以捕捉到细粒度特征, 集成多样化特征, 动态更新权重矩阵得到最终输出。

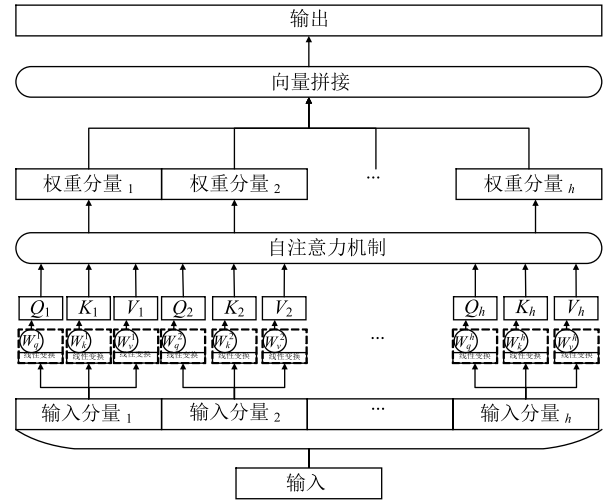


图 4 多头注意力机制结构图

Fig.4 Structure of multi-head self-attention mechanism

3 基于多头注意力机制的 LSTM-MH-SA 滑坡位移时序预测模型

3.1 LSTM-MH-SA 模型结构

滑坡诱发因素间的相互作用是高度复杂的非线性关系, 针对传统基于门控机制的深度学习模型对阶跃型滑坡突变序列预测能力不足问题, 论文结合 LSTM 和多头自注意力机制提出 LSTM-MH-SA 模型, 采用多头自注意力计算、分配自相关权重, 分层表示滑坡时序信息, 强化对滑坡阶跃变化影响较大的多因素特征, 减少异常点影响, 各隐状态加权和作为滑坡序列多因素融合特征输出, 模型结构设计如图 5 所示。图中 $y_1 \sim y_L$ 为输出的中间状态, $x_1 \sim x_L$ 为输入序列。

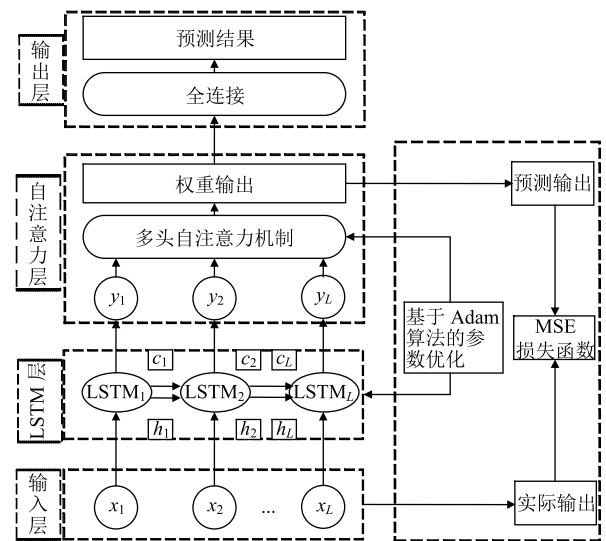


图 5 LSTM-MH-SA 结构图

Fig.5 Structure of LSTM-MH-SA

LSTM 第 t 个时间步模型隐藏层状态转移过程为

$$h_t = \text{LSTM}(h_{t-1}, c_{t-1}) \quad (13)$$

LSTM 所有时间步状态输出到多头自注意力中, 求取对应自注意力权重 a_t , 加权求和得到最终自注意力层输出 o_t :

$$a_t = \text{MH_ATT}(h_t, h_j) \quad j=1, \dots, t \quad (14)$$

$$o_t = \sum_{i=1}^t a_i h_i \quad (15)$$

如何使模型具有长程记忆能力是处理长时序累加位移数据的关键。多头自注意力为不同时刻 LSTM 隐含层的输出结果分配不同的注意力系数, 对不同阶段滑坡位移特征的重要程度进行区分学习, 解决了传统特征向量平等贡献问题, 可有效捕获不同尺度特征向量的长程相关性, 充分刻画各类诱发因素导致的滑坡局部变化特征, 提升模型记忆能力及滑坡位移突变段序列预测能力。

3.2 基于 LSTM-MH-SA 的位移预测流程

基于多头注意力机制的 LSTM-MH-SA 滑坡位移时序预测模型预测流程设计如下 (见图 6):

(1) 根据滑坡位移加法模型, 采用 VMD 分解滑坡累积位移为趋势项、周期项和随机项分量。

(2) 基于影响因子的滞后性特征, 将各位移分量和影响因子 (库水位与降雨量) 进行 DTW 相关性分析, 选择合适的影响因子并通过 VMD 分频得到模型输入序列。

(3) 对各位移分量和影响因子分量进行归一。

(4) 归一后的趋势性位移作为趋势项数据集, 周期性位移与低频影响因子组合成周期项数据集, 随机项位移与高频影响因子构成随机项数据集, 数据集分为训练集和测试集两部分。

(5) 构建 LSTM-MH-SA 模型对各位移分量进行训练和预测。用训练好的模型对各测试集进行预测, 并与实际位移进行比较。

(6) 将预测的各位移分量反归一求和得到滑坡累积位移预测结果并进行验证和分析。

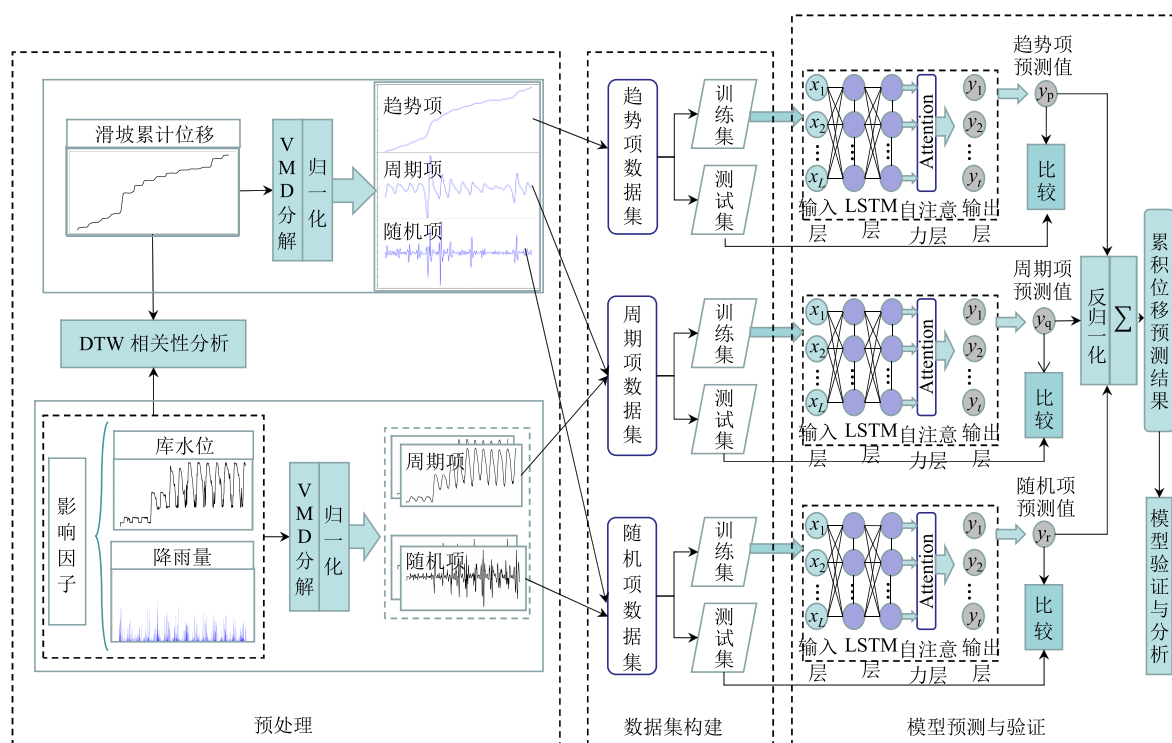


图 6 位移预测流程图

Fig.6 Flow chart of displacement prediction

4 白水河滑坡概况变形特征

4.1 滑坡工程地质概况

白水河滑坡位于长江南岸三峡库区宜昌市秭归县, 距三峡大坝约 56 km。滑坡前缘没入长江,

后缘位于岩土分界处, 两侧以基岩山脊为界。滑坡为一单斜顺向坡, 主滑方向为 NE20°, NS 向长约为 600 m, EW 向宽约为 700 m, 总体坡度约为 30°, 平均厚度约为 30 m, 体积为 $1.26 \times 10^7 \text{ m}^3$ [16]。白水河滑坡从 2003 年开始监测, 期间发生多次变形。

2003 年 6 月滑坡东部前缘产生约 300 m 长的横向裂缝, 2004 年 7 月滑坡变形导致 21 处房屋毁坏, 2005 年 8 月至 2006 年 8 月滑坡地表多处出现下沉拉式裂缝, 2007 年 6 月 30 日预警区后缘发生较大规模坍塌, 2009 年 8 月滑坡西部边界裂缝呈羽状断续扩展^[17]。

4.2 滑坡监测数据分析

白水河滑坡先后布置了 11 个 GPS 地表位移监测点, 如图 7 所示^[7]。其中监测点 ZG118 位于滑坡中段且监测时间最长, 能较好反映滑坡演变过程, 论文选取该监测点 2003 年 6 月–2016 年 12 月段滑坡累积位移数据作为试验对象, 并选择监测点 ZG93 与 XD01 作为模型适应性验证试验对象。

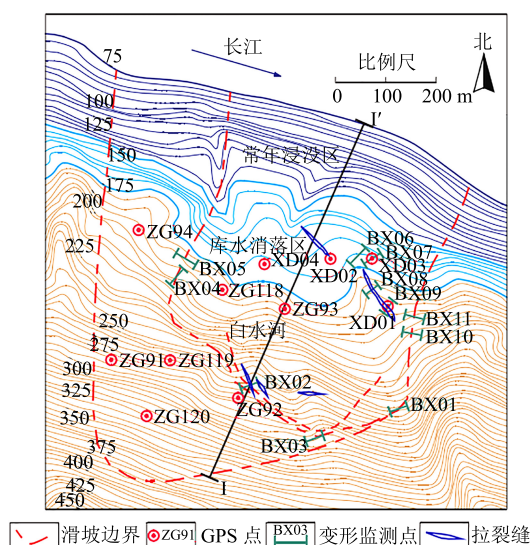


图 7 白水河滑坡 GPS 监测点布置平面图^[7]

Fig.7 Layout of GPS monitoring points in Baishuihe landslide^[7]

受滑坡地质、降雨、库水位等多因素影响, 监测点 ZG118 整体位移变化情况可分为 3 个阶段, 具体如图 8 所示。由于滑坡位移量与降雨量和库水位变化有较强的相关性, 在监测的初期 2003 年到 2006 年, 库水位较低稳定在 137 m 左右, 每年 5 月到 9 月滑坡位移量变化主要是随汛期降雨量的增大而增长, 整体变化趋势较为平缓; 在监测中期即 2006 年中期到 2009 年前期, 库水位大幅度增加至 144~155 m 附近, 库水位的突变上升对滑体应力场和岩土体结构特征影响较大, 随着汛期降雨量的增加开始人工调控库水位, 滑坡位移量变化幅度更为剧烈^[7]; 在监测后期的 2009–2013 年, 库水位长期持续在 145~175 m 高位附近波动, 趋势变化特征相对较平稳, 对滑坡地质结构影响不大, 滑坡位移量变化幅度逐渐变小。

基于上述观测, 研究进一步对滑坡累积位移各影响因素进行相关性分析。传统研究多采用灰色关联度 (gray relation analysis, 简称 GRA) 对累计处理后的月数据开展相关性分析, 由于库水位、降雨量对滑坡位移量变化的影响存在一定的时滞性, 对时效性较强的日数据采用 GRA 开展相关性分析结果不够准确。而基于动态规划策略的 DTW 算法能较好消除外界因素时滞性产生的偏差。研究基于 DTW 方法对监测点 ZG118 监测后期阶段滑坡位移量、库水位、降雨量等日数据开展相关性分析, 并与同阶段基于 GRA 月数据分析结果进行对比, 具体如表 2 所示。

由表 2 可知, 二者计算结果较为一致但 DTW 方法相关性分析结果区分度更为显著, 这是由于

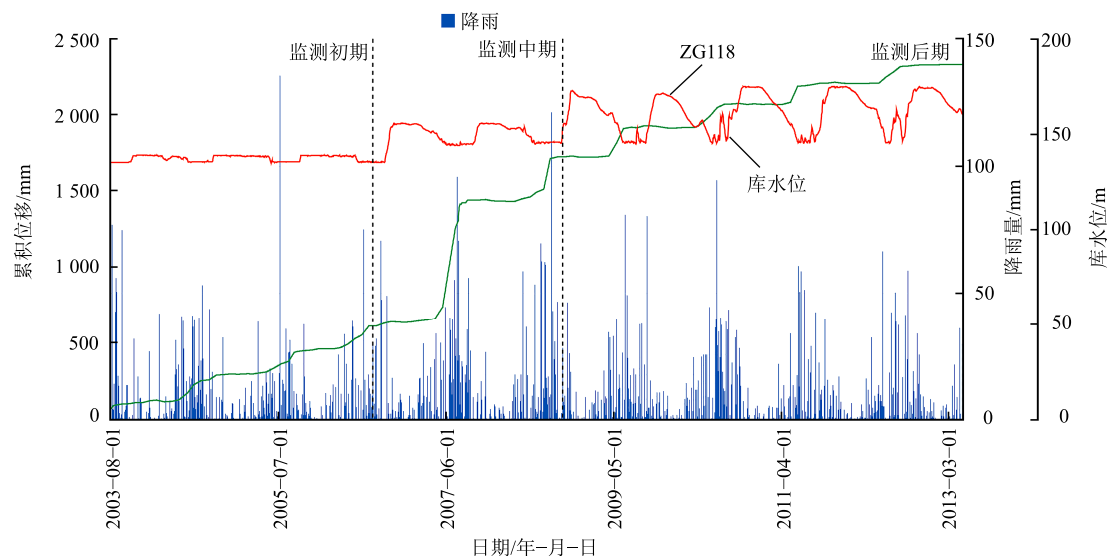


图 8 滑坡累积位移-日库水位-降雨量监测曲线

Fig.8 Monitoring curves of cumulative displacements of landslide, daily rainfall and reservoir water level

表 2 位移与影响因子关联性分析
Table 2 Correlation analysis between displacement and influence factors

位移分量类型	影响因子类型			
	库水位		降雨量	
	DTW	GRA	DTW	GRA
趋势项	0.246	0.654	0.258	0.559
周期项	0.116	0.659	0.156	0.578
随机项	0.170	0.605	0.121	0.699
总位移	0.244	0.667	0.256	0.572

注：黑体为同列对比中的最优结果，下同。

DTW 算法充分考虑了影响因子对位移的滞后性影响，而 GRA 分析的月数据只是通过放大时间尺度对数据平滑处理来消除时滞性的影响。根据相关性分析结果，库水位变化与滑坡累积位移周期项的相关性最高，说明每年汛期库水位人工调节是滑坡位移周期呈台阶变化的主要因素。降雨量与滑坡累积位移的周期项和随机项分量相关性较高，说明周期性季节降雨与非季节性的随机降雨对位移周期性和随机性变化均存在不同程度的影响。

综上所述，滑坡内部地质结构作用是引起滑坡位移量趋势性变化的主要因素，但季节性降雨衍生出的库水周期性调节使滑坡位移量变化出现周期性跃变特征，变化幅度随滑坡内部应力场、渗流场和岩土体结构适应及调整不断变小并逐渐趋于平稳，非季节性降雨及其他一些因素也会引起随机变化，众多内外因素共同作用导致滑坡累积位移总体呈现阶跃型变化特征。

5 白水河滑坡位移预测

研究对监测点 ZG118 采用提出的 LSTM-MH-SA 模型开展评估，选择 2003 年 6 月到 2013 年 12 月的日数据作为训练集，2014 年 1 月到 2016 年 12 月的日数据作为预测集，将其性能与传统 LSTM 和结合自注意力的 LSTM-SA 进行比较。

5.1 度量指标与预测模型超参数设置

采用均方根误差（root mean square error，简称 RMSE）、判定系数 R^2 （coefficient of determination）和平均绝对百分比误差（mean absolute percentage error，简称 MAPE）评价模型的性能。度量指标定义如下：

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \tilde{y}_i)^2} \quad (16)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (\tilde{y}_i - y_i)^2}{\sum_{i=1}^m (\bar{y}_i - y_i)^2} \quad (17)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^m \left| \frac{\tilde{y}_i - y_i}{y_i} \right| \quad (18)$$

式中： y_i 为 i 时刻的观测值； $\tilde{y}_i$ 为 i 时刻的预测值； $\bar{y}_i$ 为所有观测值的算术平均值； m 为样本数。RMSE 越小、 R^2 越大及 MAPE 越小，模型的预测效果越好。

研究使用 PyTorch 框架搭建模型，利用 Python 语言编写。为方便对比分析模型性能，各模型超参数设置保持一致。模型训练过程中采用 Adam 算法进行优化，学习率设置为 0.001，损失函数选择 MSE。考虑数据量和时间开销，采用网格搜索法进行参数调优，设定模型的 LSTM 层数为 2，隐藏层结点数为 64，多头自注意力多头数为 8，批处理个数为 256。具体参数设置如表 3 所示。

表 3 模型超参数设置
Table 3 Hyperparameters of model

损失函数	隐藏层节点数	学习率	多头数	隐藏层层数	批处理个数
MSE	64	0.001	8	2	256

5.2 滑坡位移量数据分解

在滑坡位移量曲线分析基础上通过多次调参，确定 VMD 中的模态参数 K 为 3，惩罚系数 α 为 300，上升步长 τ 为 0.1。VMD 算法将 ZG118 监测点数据分解为 3 部分，其中高频部分为随机项，低频部分为周期项，残余部分为趋势项，如图 9 所示。

5.3 滑坡位移预测

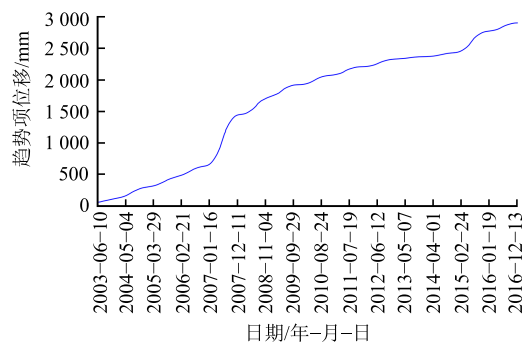
5.3.1 趋势项位移预测

趋势项位移主要受滑坡内部地质因素影响，试验选取前半个月的趋势位移量作为输入，预测后一天的趋势项位移量。算法预测结果如图 10 和表 4 所示。

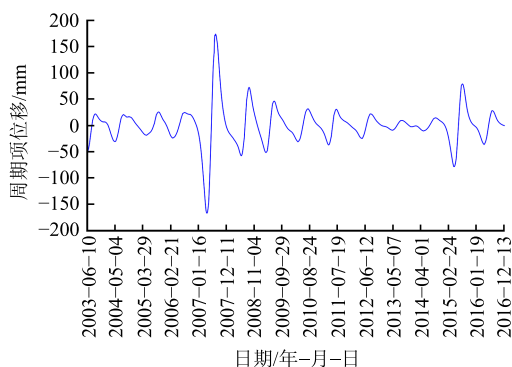
由图 10 可知，趋势性位移变化较为平稳，各模型预测趋势项均接近观测值，但自注意力机制对滑坡历史信息记忆能力具有一定的提升作用，能较好捕捉内部地质结构长期变化对趋势的影响。根据表 4，结合自注意力的精度明显比单独的 LSTM 模型高，引入多头自注意力预测效果明显优于单头自注意力预测精度。

5.3.2 周期项位移预测

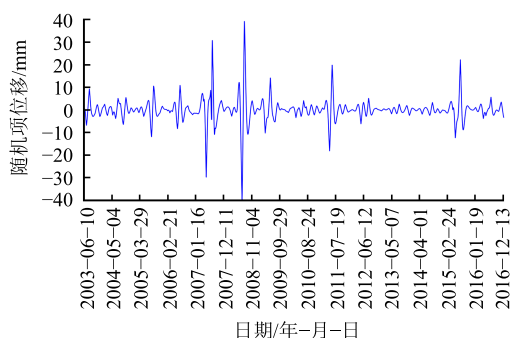
基于 4.2 节的分析，周期项位移主要受库水位人工调节与季节性降雨共同影响，而滑坡位移是两者耦合作用产生的物理结构变化与化学效应影响的结果。研究将周期性位移作为主要因子，库水位和月累积降雨量分频得到的低频分量作为辅助因子，模型预测结果如图 11 所示。



(a) 趋势项位移



(b) 周期项位移



(c) 随机项位移

图9 ZG118 监测点累积位移分解

Fig.9 Decomposition of cumulative displacement at ZG118

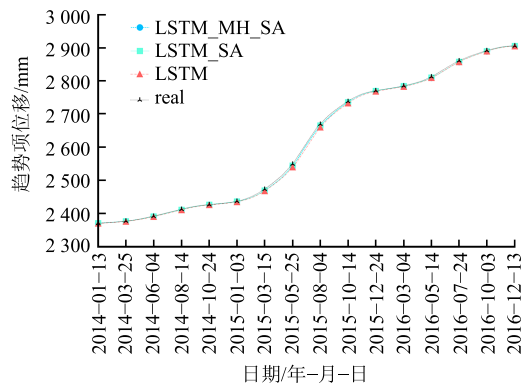


图10 趋势项位移预测

Fig.10 Prediction of trend displacement

由图 11 和表 4 可知,在滑坡位移变化较为平稳的 2014 年 1-7 月,传统 LSTM 算法预测较为准确,变化较为剧烈数据段,如 2015 年 1-12 月 LSTM 算法预测误差较大,说明门控机制对于突变数据段的特征及外在影响因素对滑坡位移的时滞性影响学习能力较弱。而结合自注意力机制的 LSTM-SA 与 LSTM-MH-SA 模型能较好刻画序列间的全局依赖关系,模型在不同的表示子空间学习相关的信息,并重点关注重要变化特征,强化对滑坡结构影响较大的周期性降雨及库水位调度影响因素的权重,故新模型在 2015 年 5-6 月、2016 年 3-6 月跃变数据段预测效果更优。整体上,结合自注意力机制的预测值与实际值吻合度高,与 LSTM 相比提升较大。较单头自注意力机制,多头自注意力机制的预测结果 MAPE 误差值略高而 RMSE 值较低说明多头自注意力对离群异常点捕捉更加准确。

5.3.3 随机项位移预测

随机项位移主要受外界一些随机因素如非季节性降雨等影响,基于相关性分析将滑坡随机性位移、库水位和前两月累积降雨量的高频随机分量作

表4 滑坡位移模型预测精度及误差

Table 4 Prediction accuracy and error of landslide displacement model

位移类型	趋势项			周期项			随机项			累积位移		
模型	RMSE/mm	R^2	MAPE	RMSE/mm	R^2	MAPE	RMSE/mm	R^2	MAPE	RMSE/mm	R^2	MAPE
LSTM	4.606	0.999 431	0.001 211	20.130	-3.09	1.926 492	5.280	-4.580	14.539 512	21.483	0.987 923	0.005 536
LSTM-SA	2.756	0.999 797	0.000 791	0.485	0.999 724	0.026 911	2.210	0.205	10.610 293	2.644	0.999 827	0.000 669
LSTM-MH-SA	2.487	0.999 835	0.000 694	0.381	0.999 830	0.036 658	1.918	0.511	11.370 371	2.601	0.999 833	0.000 638

为影响因素输入,构建模型随机项位移预测结果如图 12 所示。

由图 12 和表 4 可知高频日数据变化随机性较高,LSTM 预测结果较差,难以学习到数据随机变化特征。而结合多头自注意力的 LSTM-MH-SA 的

RMSE、 R^2 和 MAPE 分别为 1.91 mm、0.51 和 11.37,预测精度有着显著提升。

5.3.4 滑坡累积位移预测

将各分量位移相加得到滑坡的累积位移,预测结果图 13 所示。

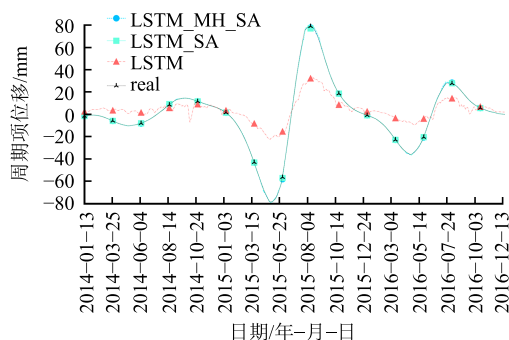


图 11 周期项位移预测

Fig.11 Prediction of periodic displacement

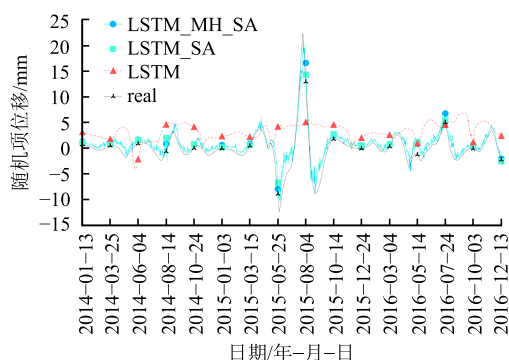


图 12 随机项位移预测

Fig.12 Prediction of stochastic displacement

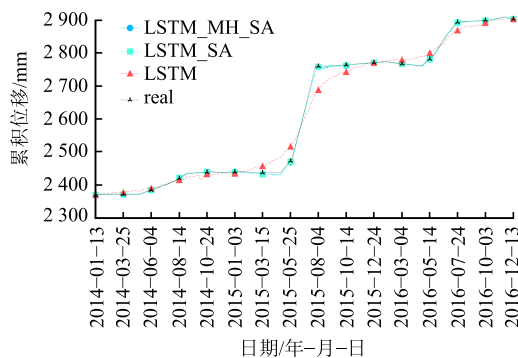


图 13 累积位移预测

Fig.13 Prediction of cumulative displacement

在滑坡产生阶跃变化时, LSTM 预测效果较差, 而结合多头自注意力的 LSTM-MH-SA 通过对滑坡多层次多尺度特征的提取, 以及对影响阶跃变化的关键因素如雨季降雨量和库水位人工调控的强化, 提升了对重要影响因子的关注和学习, 有效学习到滑坡的阶跃变化特征, 如表 4 所示。LSTM-MH-SA 的预测指标 RMSE、 R^2 和 MAPE 分别为 2.60 mm、0.999 833 和 0.000 638, 较传统 LSTM 模型有较大提升, 结合自注意力采取多头策略预测效果更好。

各模型对分频位移数据的实验结果如表 4 所示, 其中加粗部分表示各分频位移预测模型的最佳

效果。根据试验得到以下结论:

(1) 传统机器学习方法开展滑坡位移预测, 往往需要人工提取不同时间尺度的降雨量、库水位变化量等多种因子作为输入, 数据处理存在一定的随机性。而 LSTM 模型特有的遗忘门、输入门、输出门等门控机制具有较强记忆能力, 能自适应挖掘滑坡位移及各类影响因子的历史信息。自注意力机制能进一步提升 LSTM 对长程相关特征的提取能力, 在较少因子输入的情况下通过不同的表示子空间学习到不同尺度数据的特征, 捕捉序列内部的依赖关系, 预测精度更优。

(2) 传统门控机制模型 LSTM、GRU 等内部的门控组件训练随机性较强, 在变化程度剧烈的位移分量预测效果较差, 采用基准模型 LSTM 的 RMSE 为 21.48 mm, 罗祎沅等^[10]采用 GRU 在不同数据量的累积位移预测结果 RMSE 均在 20 mm 左右。而结合自注意力与多头自注意力的 LSTM-MH-SA 模型能对 LSTM 提取的所有历史状态特征进行全局学习和权重分配, 而多头自注意力更是在多个不同层面进行特征提取与融合, 提取的多样化特征对关键因素进行重点关注和深度学习, 加强捕获前期库水位骤降引起的岩土体内外物理与力学结构变化特征及降雨导致的坡面冲刷性结构和土体状态变化特征, 强化对输出影响较大参数的权重, 有效预测滑坡位移变化趋势。

5.4 模型适用性验证

为进一步验证模型在阶跃型滑坡的适用性, 选择位于白水河滑坡中部的 ZG93 监测点和滑坡前缘的 XD01 监测点为研究对象。这两个监测点较 ZG118 监测点具有不同的变形程度^[18]。试验采取相同的参数, 预测结果如表 5 所示。

从表 5 中可知, 在这两个监测点模型预测精度较高, 说明新模型针对不同变形程度和趋势的阶跃型滑坡具有较好的外推能力。

表 5 监测点 ZG93 与 XD01 累积位移预测结果
Table 5 Prediction results of cumulative displacement at ZG93 and XD01

监测点	RMSE/mm	R^2	MAPE
ZG93	2.71	0.999 822	0.000 664
XD-01	3.51	0.999 888	0.000 754

6 结 论

针对阶跃型滑坡具有的局部突变时序数据特征, 论文提出结合 LSTM 与多头自注意力机制的 LSTM-MH-SA 模型。采用该模型对三峡库区白水河

滑坡位移开展预测研究, 试验结果表明预测精度较传统动态预测模型有较大提升, 为阶跃型滑坡预测提供新的解决思路。新模型主要有以下特点:

(1) 结合自注意力机制解决了 LSTM 门控机制难以学习到局部阶跃数据变化趋势的问题, 周期位移与随机项预测结果表明多头注意力机制能够提升累积位移的预测精度。

(2) 模型通过对 LSTM 所有时间步的隐藏层状态进行自相关权重计算提升了模型的记忆能力, 能充分描述不同诱发因素导致的滑坡局部变化特征, 解决数据时效难题。

(3) 通过多头自注意力对数据的多尺度信息提取与集成, 有效缓解传统模型训练存在的过拟合问题。

由于自注意力机制各时间步都需额外进行权重计算, 模型的训练时间开销较大。后期将在稀疏自注意力机制方面考虑对 q 、 k 进行稀疏性选择, 以降低时间复杂度。

参考文献

- [1] 张凯, 张科, 保瑞, 等. 基于优化经验模态分解和聚类分析的滑坡位移智能预测研究[J]. 岩土力学, 2021, 42(1): 211-223.
ZHANG Kai, ZHANG Ke, BAO Rui, et al. Intelligent prediction of landslide displacements based on optimized empirical mode decomposition and K-mean clustering[J]. **Rock and Soil Mechanics**, 2021, 42(1): 211-223.
- [2] 邓冬梅, 梁烨, 王亮清, 等. 基于集合经验模态分解与支持向量机回归的位移预测方法: 以三峡库区滑坡为例[J]. 岩土力学, 2017, 38(12): 3660-3669.
DENG Dong-mei, LIANG Ye, WANG Liang-qing, et al. Displacement prediction method based on ensemble empirical mode decomposition and support vector machine regression-a case of landslides in Three Gorges Reservoir area[J]. **Rock and Soil Mechanics**, 2017, 38(12): 3660-3669.
- [3] 黄发明, 殷坤龙, 杨背背, 等. 基于时间序列分解和多变量混沌模型的滑坡阶跃式位移预测[J]. 地球科学, 2018, 43(3): 887-898.
HUANG Fa-ming, YIN Kun-long, YANG Bei-bei, et al. Step-like displacement prediction of landslide based on time series decomposition and multivariate chaotic model[J]. **Earth science**, 2018, 43(3): 887-898.
- [4] 刘艺梁, 殷坤龙, 汪洋, 等. 基于经验模态分解和神经网络的滑坡变形预测研究[J]. 安全与环境工程, 2013, 20(4): 14-17.
LIU Yi-liang, YIN Kun-long, WANG Yang, et al. Study of landslide deformation prediction based on EMD and neural network[J]. **Safety and Environmental Engineering**, 2013, 20(4): 14-17.
- [5] 张俊, 殷坤龙, 王佳佳, 等. 基于时间序列与 PSO-SVR 耦合模型的白水河滑坡位移预测研究[J]. 岩石力学与工程学报, 2015, 34(2): 382-391.
ZHANG Jun, YIN Kun-long, WANG Jia-jia, et al. Displacement prediction of Baishuihe landslide based on time series and PSO-SVR model[J]. **Chinese Journal of Rock Mechanics and Engineering**, 2015, 34(2): 382-391.
- [6] ZHOU C, YIN K, CAO Y, et al. Application of time series analysis and PSO-SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China[J]. **Engineering Geology**, 2016, 204: 108-120.
- [7] 李麟玮, 吴益平, 苗发盛, 等. 基于变分模态分解与 GWO-MIC-SVR 模型的滑坡位移预测研究[J]. 岩石力学与工程学报, 2018, 37(6): 1395-1406.
LI Lin-wei, WU Yi-ping, MIAO Fa-sheng, et al. Displacement prediction of landslides based on variational mode decomposition and GWO-MIC-SVR model[J]. **Chinese Journal of Rock Mechanics and Engineering**, 2018, 37(6): 1395-1406.
- [8] 檀梦蛟, 殷坤龙, 郭子正, 等. 基于 CEEMDAN 理论和 PSO-ELM 模型的滑坡位移预测[J]. 地质科技情报, 2019, 38(6): 165-175.
TAN Meng-jiao, YIN Kun-long, GUO Zi-zheng, et al. Landslide displacement prediction based on CEEMDAN method and particle swarm optimized-extreme learning machine model[J]. **Geological Science and Technology Information**, 2019, 38(6): 165-175.
- [9] 杨背背, 殷坤龙, 杜娟. 基于时间序列与长短时记忆网络的滑坡位移动态预测模型[J]. 岩石力学与工程学报, 2018, 37(10): 2334-2343.
YANG Bei-bei, YIN Kun-long, DU Juan. A model for predicting landslide displacement based on time series and long and short term memory neural network[J]. **Chinese Journal of Rock Mechanics and Engineering**, 2018, 37(10): 2334-2343.
- [10] 罗祎沅, 蒋亚楠, 许强, 等. 最优分解模态和 GRU 模型的库岸滑坡位移预测研究[J/OL]. 武汉大学学报(信息科学版), 2021, DOI: 10.13203/j.whugis20200610.
LUO Hui-yuan, JIANG Ya-nan, XU Qiang, et al. Study on displacement prediction of reservoir bank landslide based on optimal decomposition mode and GRU model[J]. **Geomatics and Information Science of Wuhan University**, 2021, DOI: 10.13203/j.whugis20200610.