

基于 ARIMA-LSTM-XGBoost 组合的矿山排土场沉降预测模型研究

高梓皓^{1,a,*}, 智浩森^{1,b}, 黄树林^{1,c}, 张世杰^{1,d}, 纪舒南^{1,e}

(1.中国辽宁省葫芦岛市兴城, 125000, 辽宁工程技术大学安全科学与工程学院, 安全工程;

* a. 2085841721@qq.com; b. 1501721961@qq.com; c. 1765041333@qq.com; d. 2307471359@qq.com;
e. 2639147655@qq.com)

摘要: 为确保矿山排土场边坡安全稳定, 提出使用 GNSS 技术对其沉降量进行监测研究。以华新水泥矿山排土场为例, 基于 2022 年 1 月 1 日 0 点至 2022 年 4 月 31 日 23 点间每小时采集的历史数据, 通过箱线图和累积沉降量时间序列关系图进行数据分析。考虑到单一模型不足、预测精度等问题, 选择结合自回归综合移动平均模型 (ARIMA)、长短时记忆网络模型 (LSTM) 和极端梯度提升模型 (XGBoost), 建立 ARIMA-LSTM-XGBoost 边坡沉降预测模型, 将组合模型预测结果与实际值及单一模型进行对比, 并以绝对相对误差 (ARE) 和均方根误差 (RMSE) 作为评价指标。研究结果表明: 根据历史监测数据, 边坡 3 相较于边坡 1 与边坡 2 更为稳定; 三个边坡测试样本的 ARIMA-LSTM-XGBoost 组合预测最大绝对相对误差为 4.58%、6.45%、7.55% 其误差远小于其他 4 种对比模型, 研究结果可为矿山工程技术人员对矿场边坡沉降预测提供参考。

关键词: 稳定性; 边坡沉降; GNSS 技术; ARIMA 模型; LSTM 模型; XGBoost 模型; 误差

引言

排土场通常位于山坡或沟谷, 是露天矿开采的必然产物。此类区域用于储存和处理矿山开采过程中产生的大量废弃物、土石和矿石[1-2]。一旦排土场发生不稳定, 可能会导致严重的安全问题和环境影响, 对矿山运营和周边生态系统构成严重威胁, 进而影响矿山生产的安全和顺利进行。因此, 对于矿山排土场边坡沉降预测问题的研究具有较重要的实际意义[3-8]。

近些年来, 众多学者对矿山边坡沉降预测问题展开深入研究。一种结合时间序列分析理论来建立小波包-时间序列预测模型, 该模型用于露天矿边坡变形分析预测[9]。Luo et al.[10]提出了一种基于多输出相关向量机 (MRVM) 的边坡变形预测新模型, 基于边坡变形数据展开研究。上述模型或者数学方法可以较好地预测边坡沉降, 但存在一定的缺陷: 模型或者数学方法计算复杂, 推广性较差; 部分只考虑使用单一模型, 存在模型本身所固有的不足; 对于在处理时间序列预测精度上还亟待提高。

综上所述, ARIMA 模型适用于处理线性关系和季节性趋势, LSTM 模型能够捕捉序列数据中的长期依赖关系, XGBoost 则擅长处理非线性关系和复杂的特征空间, 并且上述三种模型已经在各个领域得到广泛应用, 具有较好的推广性。通过将这三种模型结合起来, 可以综合利用它们在不同方面的优势, 提高预测性能。基于 GNSS 技术对华新水泥矿山排土场历史沉降进行监测, 得到历史沉降数据之后, 选择使用数学方法对其进行分析, 并使用组合预测模型对其展开研究, 将组合预测的预测结果与单一模型进行比较分析, 验证其精确性。

1 模型原理介绍

ARIMA(p,d,q) 模型是一种时间序列分析和预测的方法该模型结合了自回归 (AR)、差分 (I) 和移动平均 (MA) 这三个组成部分, ARIMA (p,d,q) 模型可以根据矿山排土场边坡历史沉降数据来预测未来沉降值, 模型中的参数 q 表示自回归系数; p 表示偏自回归系数; d 表示差分系数。建模基本流程如下: 首先对时间序

列平稳性进行检验，利用差分将非平稳序列转化成平稳序列，然后确定差分参数（d），接着通过观察偏自相关（PACF）以及自相关图（ACF）来确定自回归参数（p）和移动平均参数（q）。

1.1 LSTM 模型

LSTM（长短期记忆网络）是一种深度学习中的循环神经网络（RNN）的变体，专门设计用来解决长序列依赖性问题，其原理结构如下图 1 所示。由于矿山排土场边坡沉降现象在未来一段时间内一直存在，选择 LSTM 模型能够捕捉和记忆长期依赖关系的能力，这对于矿场边坡沉降预测问题较为适用。

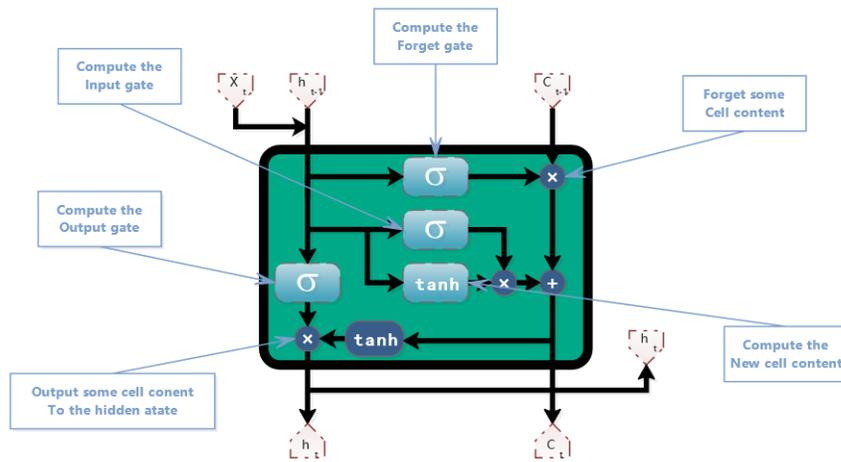


图 1 LSTM 原理结构图

XGBoost 模型是一种基于 boosting 集成学习的预测模型，可以用于矿场边坡沉降预测问题[。基本思路为：由多个弱分类器集成建立强分类器，优化经验损失函数后，在反复迭代中不断对损失函数的负梯度进行拟合，然后使用线性搜索方法，以此生成最优的弱学习器。

ARIMA-LSTM-XGBoost 预测模型在矿山边坡沉降量预测领域中的应用具有显著的创新性。该组合模型融合了时间序列分析、深度学习和机器学习技术的混合模型，为矿山排土场边坡沉降预测提供了强大的工具。基于上述三种算法原理的分析，为了提升沉降量的预测精度，采用误差修正和基于权重分配搭配的方式对 ARIMA 模型、LSTM 模型以及 XGBoost 模型进行组合，将其应用在矿山边坡沉降预测领域。图 2 为其组合后的 ARIMA-LSTM-XGBoost 预测模型流程图。

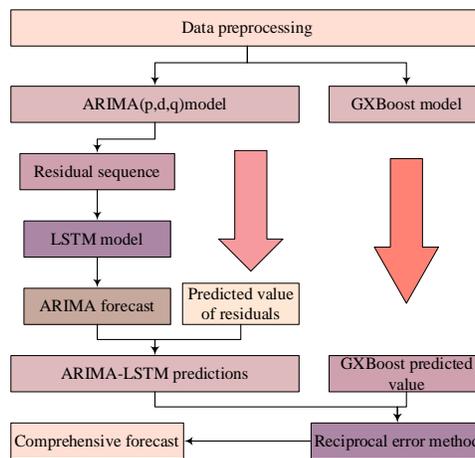


图 2 ARIMA-LSTM-GXBoost 预测模型流程图

2 数据收集与分析

以某矿山排土场 3 个不同位置的边坡为研究对象，利用 GNSS 技术对这 3 个边坡的沉降展开监测。GNSS 代表全球导航卫星系统，是一种利用卫星提供全球范围内定位、导航和定时服务的技术系统。利用 GNSS 提供实时的地理位置信息，使得矿场边坡沉降情况可以及时监测，监测的历史数据包括 3 个边坡位置 2022 年 1 月 1 日 0 点-2022 年 4 月 30 日 23 点的每个小时的沉降情况，部分数据见下表 1：

表 1 监测的历史沉降数据

时间	边坡 1	边坡 2	边坡 3
2022-01-01 00 点	1.03	0.91	0.51
2022-01-01 01 点	0.15	1.21	0.07
2022-01-01 02 点	-0.22	1.06	-0.11
2022-01-01 03 点	-0.20	-0.09	-0.10
.....
2022-04-30 20 点	0.60	0.27	0.30
2022-04-30 21 点	0.74	0.32	0.36
2022-04-30 22 点	0.69	0.43	0.34
2022-04-30 23 点	0.67	0.66	0.33

根据华新水泥矿山排土场边坡沉降的历史数据可以对这 3 个边坡沉降进行分析，分析过程如下：首先利用历史数据通过 Origin 软件绘制出 3 个边坡的箱线图，如下图 3 所示：

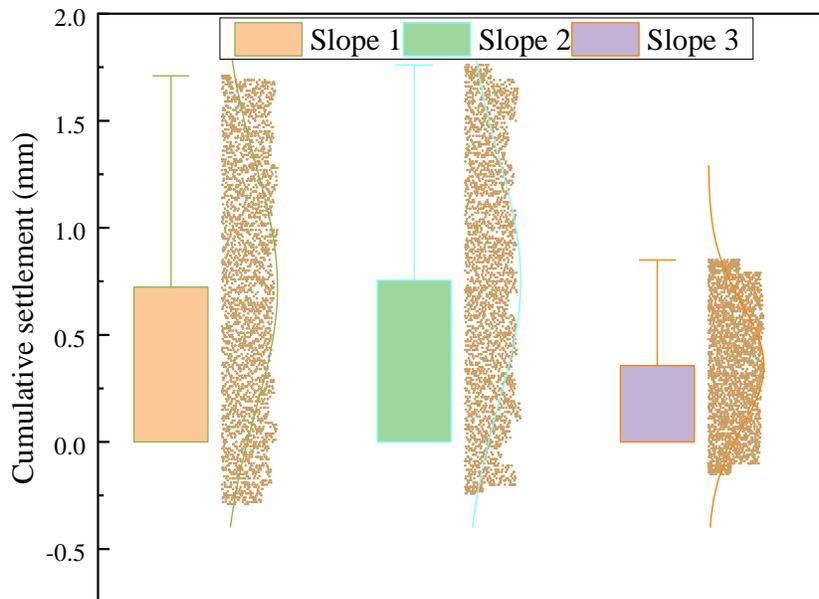


图 3 3 个边坡线箱图

由上图 3 分析可知，边坡 1、边坡 2 的沉降数据分布在 $[-0.35, 1.75]$ 之间，而边坡 3 的沉降数据只分布在 $[-0.25, 0.85]$ 之间，并且边坡 1 与边坡 2 的每小时沉降量几乎在 $[0.15, 1.45]$ 之间。对比边坡 2 发现其几乎也在 $[0.15, 1.45]$ 之间，而边坡 3 的沉降却几乎在 $[0.05, 0.75]$ 之间。

接着通过绘制出 3 个边坡的累积沉降量与时间的关系以边坡累积沉降速度，以此来初步判断 3 个边坡的沉降速度，如下图 4 所示：

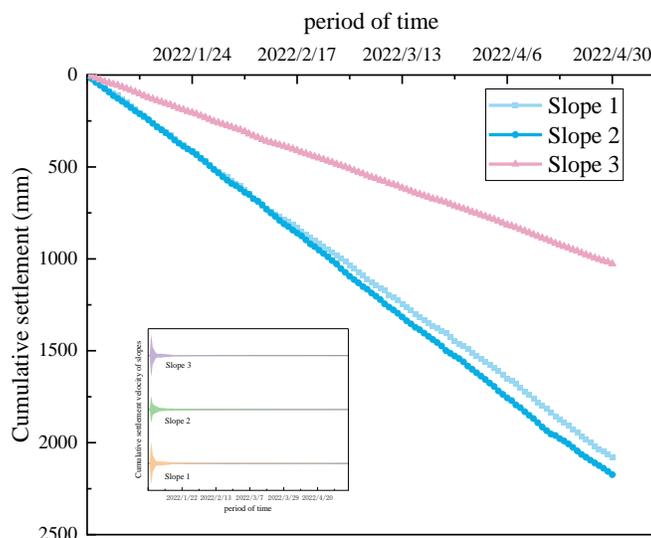


图 4 边坡累积沉降量与时间关系图

观察图 4 可知,随着时间的推移,这 3 个边坡的累积沉降量不断增加,其中边坡 2 的斜率最为陡峭,沉降速度最快;而边坡 1 的沉降速度仅次于边坡 2,对于边坡 3,其斜率相比于边坡 1 与 2 较为平缓,沉降速度较慢,其安全性高于边坡 1 与边坡 2,预警状态小于其他两个边坡;同时对比边坡累积沉降速度,其边坡 1 和边坡 3 的起始累积沉降速度要大于边坡 2 的,并随着时间推移变得平缓,危险预警系数逐渐升高,处于较为危险的状态。

3 预测结果分析

建立 ARIMA-LSTM 预测模型和 GXBoost 模型后,以 2022 年 1 月 1 日 0 点-3 月 31 日 23 点的 3 个边山排土场边坡沉降数据作为训练集,以 2022 年 4 月 1 日 0 点-4 月 30 日 23 点的数据作为测试集,得到两个模型的预测值和预测误差后对两个模型进行加权组合,得到组合模型的预测值,并计算其预测误差,预测误差使用绝对相对误差来表示。绝对相对误差(ARE)是用来衡量某个估计值与真实值之间差异的一种度量方法,其公式如下式所示:

$$ARE = \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\%$$

式中: y_i 表示沉降量实际值, \hat{y}_i 表示沉降量预测值。

基于上述分析,利用 MATLAB 软件构件上述 ARIMA-LSTM-GXBoost 组合模型,并将预测结果与 ARIMA、LSTM、GXBoost、ARIMA-LSTM 模型的预测值以及实际值进行对比,同时对比其误差的大小,考虑到对比图形的可视化,选择测试集中的任意 8 组数据,并标记为样本 1-8 号,用于模型的预测结果及误差对比。各个模型的预测结果及误差对比如下图 5、6、7 所示:

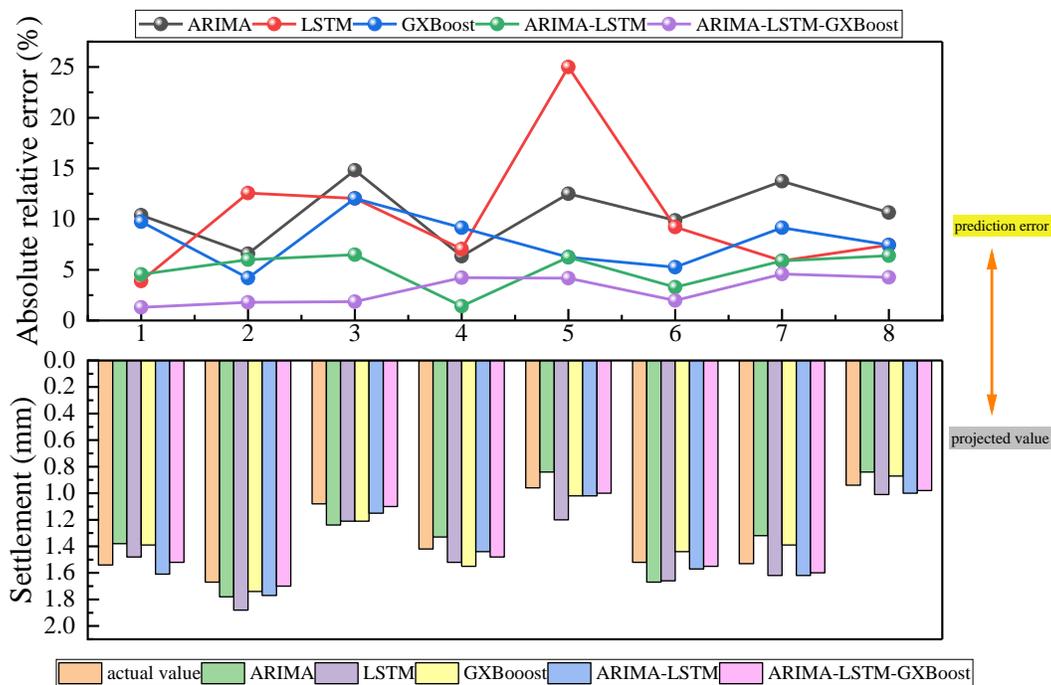


图5 边坡1不同模型的预测结果与预测误差对比

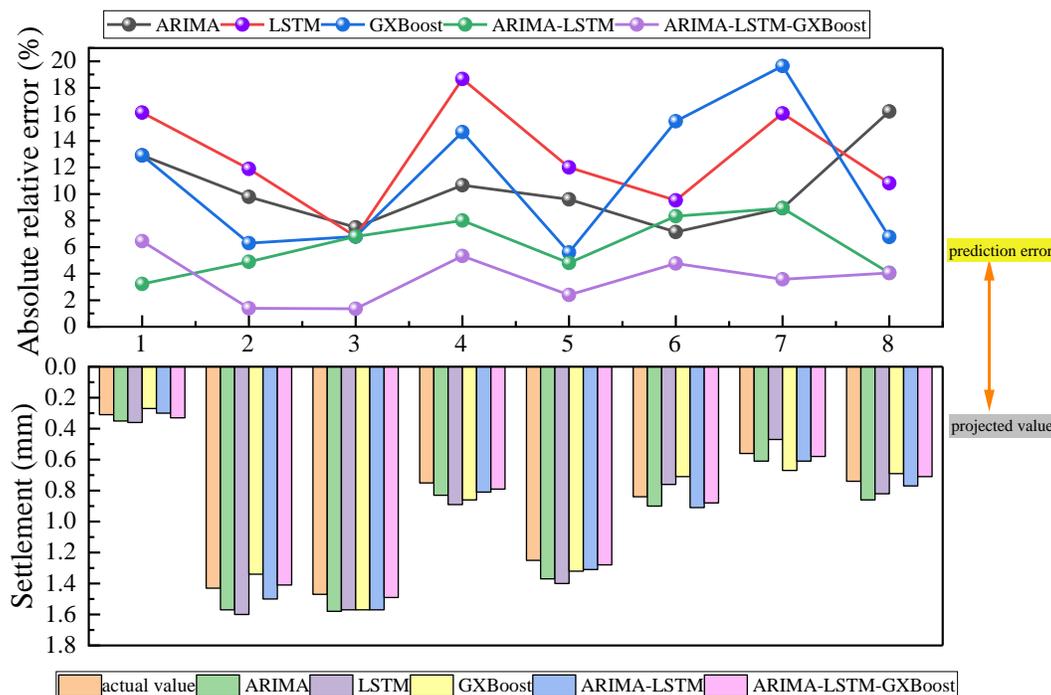


图6 边坡2不同模型的预测结果与预测误差对比

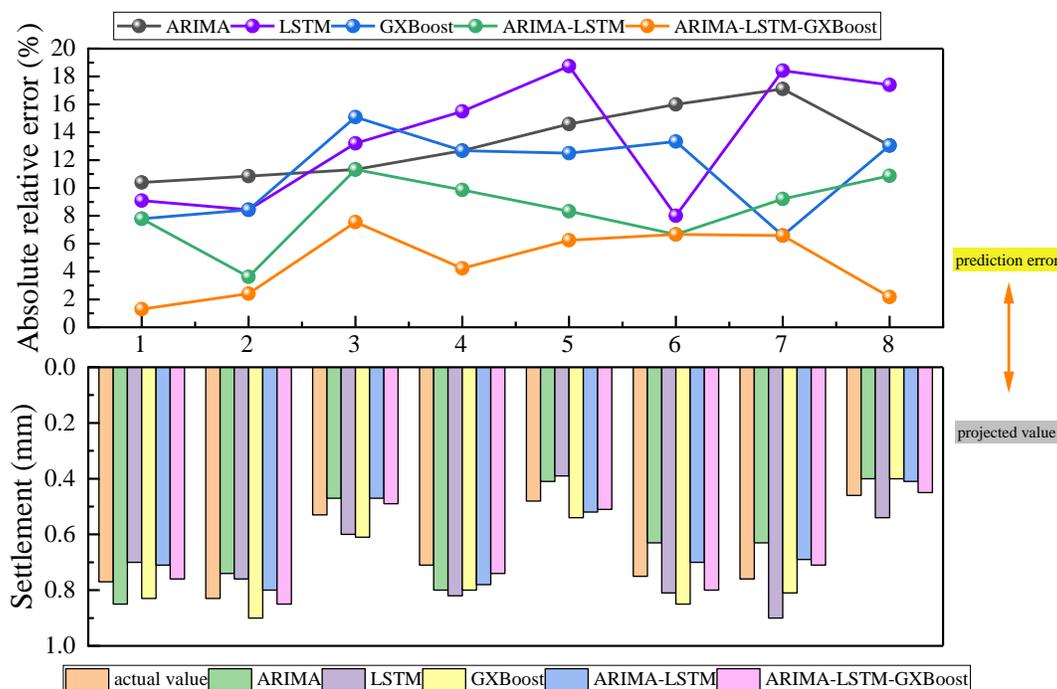


图 7 边坡 3 不同模型的预测结果与预测误差对比

图 5、6、7 展示了 3 个边坡垂直沉降上不同预测模型预测结果以及预测精度对比情况，发现 ARIMA-LSTM-GXBoost 组合预测模型的预测结果与实际值最为贴近，说明该组合预测模型适用于矿山排土场边坡沉降预测问题。同时纵向对比其他预测模型，其预测效果较好。

由此分析发现：建立的 ARIMA-LSTM-GXBoost 组合模型可以帮助矿山管理者评估边坡稳定性，并采取必要的措施来确保矿山的安全运营。如果沉降量过大可能会导致边坡失稳，从而引发坍塌和滑坡等灾害，威胁到人员安全和设施的完整性。准确预测有助于采取措施减少这些负面影响。

4 结论

1) 利用 GNSS 技术对某矿山排土场边坡沉降进行监测，使用数学方法对监测的历史数据展开分析，发现边坡 3 的安全状态最好，其次是边坡 1，而边坡 3 的危险系数最高，需要重点关注并采取防护措施。

2) 在确定各个模型的参数之后，采用训练好的 ARIMA-LSTM-GXBoost 组合模型来对测试样本进行预测，发现该组合模型更符合实际，有利于预防边坡失稳从而引发坍塌和滑坡等灾害。

3) 建立的组合预测模型在 3 个边坡上的最大绝对相对误差为 4.58%、6.45%、7.55%，更准确的预测结果有助于为工程技术人员管理矿山的安全运营提高借鉴。

参考文献

- [1] Wang Wenfei, YANG Zhiqian, MENG Xiangrui et al. Slope stability analysis of mine drainage field based on the combination of FLAC^{2D}, GPS monitoring and manual patrolling [J]. Nonferrous Metals (Mining Section), 2023, 75(6): 80-87.
- [2] Cai Xueding, Gao Shisong, Meng Linbo. Analysis on influencing factors of slope stability in dump site of open-pit mine based on SPSS [J]. Journal of Safety Science and Technology. 2023, 19(S1): 19-24.
- [3] Qi Q, Yue X, Duo X, et al. Spatial prediction of soil organic carbon in coal mining subsidence areas based on RBF neural network [J]. International Journal of Coal Science & Technology, 2023, 10(1): 30.
- [4] Wu X, Feng Z, Liu Y, et al. Enhanced safety prediction of vault settlement in urban tunnels using the pair-copula and Bayesian network [J]. Applied Soft Computing, 2023, 132: 109711.

- [5] Zhang X, Li W, Zhang X, et al. Application of grey feed forward back propagation-neural network model based on wavelet denoising to predict the residual settlement of goafs[J]. Plos one, 2023, 18(5): e0281471.
- [6] Zhong Z, Xu Y, Wang N, et al. Environmental characteristics and unified failure mode classification system for mining landslides in the karst mountainous areas of southwestern China[J]. Carbonates and Evaporites, 2023, 38(1): 2.
- [7] Qiao C, Guo Y, Li C. Study on rock burst prediction of deep buried tunnel based on cusp catastrophe theory[J]. Geotechnical and Geological Engineering, 2021, 39: 1101-1115.
- [8] Moayedi H, Rezaei A. The feasibility of PSO-ANFIS in estimating bearing capacity of strip foundations rested on cohesionless slope[J]. Neural Computing and Applications, 2021, 33: 4165-4177.
- [9] Zhao Limin, GAO Ang, QI Yongbo et al. A model of settlement prediction of open-pit mine slope based on time series analysis [J]. Engineering of Surveying and Mapping, 2017,26(9):46-50.
- [10] Luo Yiyong, Yao Yibin, Wang Ya. Multi Output Slope Deformation Model Based on Multi-Output Relevance Vector Machine [J]. Journal of Geodesy and Geodynamics, 2017,37(12):1239-1242.