Uncertainty elevation of landslide displacement prediction based on LSTM and Mixture Density Network

Huafu Pei\textsuperscript{1}, Fanhua Meng\textsuperscript{2}, Lin Wang\textsuperscript{3}

\textsuperscript{1,2} Department of Geotechnical Engineering, Dalian University of Technology
\textsuperscript{3} Chuo Kaishatsu Corporation, Japan

\textsuperscript{1} huafupei@dlut.edu.cn
\textsuperscript{2} mfh924625@mail.dlut.edu.cn

Abstract. Considering monitoring noise and complexity of landslide movement, quantify the uncertainty of landslide displacement prediction is crucial and challenging. Traditional data-driven models such as long short-term memory networks (LSTM), support vector regression, and extreme learning machines, etc., give the predicted displacement without considering the uncertainty of the predictions. Moreover, the loss function of the data-driven model is mainly taken as mean square error, which may lead to a worse performance when the training data follows non-normal distribution and thus reduces the robustness of the model. This study tends to propose a novel hybrid model based on LSTM and mixture density network to quantify each data point’s probability density distribution. By introducing the mixture density network and the maximum likelihood loss function, this model can get rid of the limitation that the data needs to obey the normal distribution. The mixed probability density parameters of each data point were predicted accurately due to the dynamic learning process in time series by LSTM. Moreover, we introduce the ensemble prediction to consider the uncertainty of model parameters. The performance of the model was validated based on a typical landslide in the Three Gorges Reservoir Area (TGRA), the Baishuihe landslide. Application results demonstrate that the proposed model provides accurate mean predictions and reasonable confidence displacement intervals.

1. Introduction

As the fourth global killer, landslides have seriously threatened human life and property after floods, storms, and earthquakes [1]. China has the most recorded landslides in Asia, with a total area of 173.52 × 10\textsuperscript{4} km\textsuperscript{2}, accounting for 18.10\% of the total land area in China [2]. Due to the complex geology environment and variant external trigger factors, landslide movement is a dynamic and mutable process, which leads to a reliable and accurate landslide displacement prediction necessary for an early warning system. Recently, many data-driven models have been introduced into landslide displacement prediction successfully [3–6], however, these models only provide the point prediction, which can be treated as the mean prediction from the perspective of probability. Due to the limited monitoring data, the results of landslide displacement prediction based on monitoring data contain the uncertainty of data and model. Quantitative evaluation of prediction uncertainty can provide more information for landslide early warning. Moreover, although some hybrid models have been proposed to predict displacement intervals to elevate prediction uncertainty [7,8], the direct prediction of displacement intervals causes the evaluation results to be contrary to reality, for example, when the displacement...
increases sharply (with less monitoring data, i.e., small probability event), the elevated uncertainty is less (narrower intervals). To address these problems, we introduce a mixture density network (MDN) to model each data point’s probability density and the prediction uncertainty can be elevated by Monte Carlo sampling. Meanwhile, we adopt ensemble prediction results to consider the model’s uncertainty caused by the initial values of the model’s parameters. Considering that MDN is a fully connected neural network, which can not learn the sequence data well, we introduce LSTM to consider the sequence correlation.

2. Methodology
The methodology adopted in this paper is divided into four parts and the main integration process is described as a flowchart (Figure. 1). **Step 1:** The cumulative displacement is decomposed into trend and periodic terms by moving average algorithm. **Step 2:** Trend displacement is clustered by the k-means algorithm, and the last cluster is fitted by cubic polynomial for trend prediction. **Step 3:** Periodic displacement’s probability density is predicted by LSTM-MDN considering reservoir water level and precipitation trigger factors, and 100 hybrid models with different initial parameters are trained based on the same dataset and the average results are calculated. **Step 4:** The mean and confidence intervals of cumulative displacement are obtained by combining the trend displacement with the sampled periodic displacement.

Figure. 1 Flowchart of the proposed models for cumulative displacement prediction
2.1. LSTM
LSTM is a typical recurrent neural network (RNN) that contains recurrent connections between the different time steps in one unit in a hidden layer to pass information from one step (time step $t-1$) to the next one (time step $t$). To avoid the vanishing gradient of the recurrent connections when the network processes long time series, LSTM introduces three gate units (“input gate”, “forget gate” and “output gate”) to control the flow of series information. The input gate decides how much information should be fed in the current time step. The forget gate controls whether the information from the previous time step is remembered or forgotten. The output gate decides how much information should be passed into the next time step or to the final results [9,10]. The mathematical formulas of the hidden layer are shown as follows:

$$A_t = \sigma(W_{ix}x_t + W_{hh}h_{t-1} + W_{cc}c_{t-1} + b_i), A = (i, o, f)$$ (1)
$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$ (2)
$$h_t = o_t \tanh(c_t)$$ (3)

where $i_t, o_t, f_t$ and $c_t$ are the values of the input gate, output gate, forget gate, and the memory cell in LSTM block; $b_i, b_o, b_f$ and $b_c$ are corresponding bias values; $\sigma$ represents the sigmoid function; $W_{ix}, W_h$ and $W_c$ represent the weights between input and hidden nodes, hidden and memory cell, and memory cell and outputs, respectively.

2.2. MDN
The MDN is a conventional neural network combined with a mixture density model, and it can in principle represents arbitrary conditional probability distributions. [3] demonstrates that the network function is given by the conditional average of the target data, conditioned on the input data, which is formulated as Eq. (4).

$$f_k(x, w^*) = \int y_k p(y_k | x) dy$$ (4)

where $f_k(x, w^*)$ represents the network prediction of $k$th data point; $y_k$ represents the values of the $k$th target data; $p(y_k | x)$ represents the probability density of $y_k$ conditioned on the input data. The core of this model is to replace the conditional probability density in Eq. (4) with a mixture model in Eqs. (5), (6), which has the flexibility to model completely general distribution functions.

$$p(y | x) = \sum_{i=1}^{m} \alpha_i(x)\phi_i(t | x), \quad \sum_{i=1}^{m} \alpha_i(x) = 1$$ (5)
$$\phi_i(y | x) = \frac{1}{(2\pi)^{c/2}\sigma_i(x)} \exp \left\{- \frac{||y - \mu_i(x)||^2}{2\sigma_i(x)^2} \right\}$$ (6)

where $m$ is the number of components in the mixture and $\alpha_i(x)$ is the mixing coefficient; $\phi_i(y | x)$ is the Gaussian kernel function with the centre $\mu_i(x)$ and variance $\sigma_i(x)$; $c$ represents the dimension of $y$. The dimension of MDN output for one data point is $3m$, including $m \alpha_i(x), m \mu_i(x), and m \sigma_i(x).$

In this study, the mixing coefficients are obtained by the “softmax” function, and variance $\sigma_i(x)$ is obtained by the “soft plus” function (Eq. (6)) for the non-negative requirement.

$$\alpha_i = \frac{\exp(\alpha_i)}{\sum_{j=1}^{m} \exp(\alpha_j)}, \quad \sigma_i(x) = \ln \left[ \exp(\sigma_i(x)) + 1 \right]$$ (7)
Due to the probability density distributions of data points provided by MDN, the corresponding loss function should be defined as the maximum joint density \( p(y, x) \) (or logarithmic form \( \ln p(y, x) \)).

\[
\max \{ \ln p(t, x) \} = \max \left\{ \sum_{i=1}^{n} \ln \left( p(t|x_i) p(x) \right) \right\} = \max \left\{ \sum_{i=1}^{n} \ln \sum_{i=1}^{m} \alpha_k(x_i) \varphi(t|x_i) + \sum_{i=1}^{n} \ln p(x) \right\} = \max \left\{ \sum_{i=1}^{n} \ln \sum_{i=1}^{m} \alpha_k(x_i) \varphi(t|x_i) \right\} + \text{Const}
\]

(8)

3. Baishuihe landslide case study

The Baishuihe landslide is located in the town of Zigui, Hubei Province, China, on the south bank of the Yangtze River [11]. It is a fan-shaped, retrogressive landslide with a coverage area of 0.42 km\(^2\), a maximum length of 780 m, a width of 430 m, and an average thickness of 30 m. The landslide slopes in the direction of N 20° E and the estimated volume is 1260×10\(^4\) m\(^3\). According to the site investigation and monitoring data, the Baishuihe landslide can be divided into two blocks, i.e., an active block A and a relatively stable block B. The sliding body mainly consists of cataclastic rock, silty mudstone, and gravelly soil. Two sliding surfaces are observed at different depths (Figure 2). The lithology of bedrock mainly contains Jurassic siltstone, silty mudstones, and quartz sandstone, with dip directions of 15° and dip angles of 36° [12]. Eleven GPS stations were installed on the landslide, and the ZG118 monitoring station, which is located in the center of the landslide and on block A, is adopted to represent the landslide’s behavior. In this study, we aim to predict one-year cumulative displacement, thus, the last twelve monitoring data points are excluded as the testing dataset.

4. Results analysis

4.1. Displacement decomposition and trend displacement prediction

The moving average algorithm was adopted to decompose the ZG118 cumulative displacement into trend and periodic displacement. Let original sequence be \( D(t) = [d_1, d_2, \ldots, d_n] \), then the trend displacement was calculated as follows:

![Fig. 2 Geological section of the Baishuihe landslide](image-url)
\( \Phi(t) = \frac{d_t + d_{t-1} + \cdots + d_{t+n}}{k}, (t = k, k+1, \ldots, n) \) 

(9)

where \( n \) is the number of monitoring data and \( k \) is the moving cycle and was set to 12. The k-means algorithm was adopted to cluster trend displacement for a better prediction. The decomposition and cluster results along with the fitting and predictive results of ZG118 are shown in Figure. 3, and the corresponding regression coefficients are listed in Table 1.

![Figure. 3 Displacement decomposition, trend cluster, and trend prediction results of ZG118](image)

| Period            | a      | b      | c  | d     | Correlation coefficient \((R^2)\) |
|-------------------|--------|--------|----|-------|---------------------------------|
| July-2004 to Oct-2007 | 1.63e-2 | -7.67e-1 | 23.86 | 96.58 | 0.991                           |
| Nov-2007 to Oct-2009 | 1.01e-1 | -4.72  | 93.63 | 8.96e2 | 0.997                           |
| Nov-2009 to Dec-2012 | 1.02e-4 | -6.44e-2 | 14.61 | 1.81e3 | 0.999                           |

4.2. Periodic displacement prediction

According to recent researches [7,9,12,13], reservoir water level and precipitation are two main factors that influence periodic displacement. Considering that LSTM can learn the correlation between the sequence data, the reservoir water level and precipitation datasets corresponding to the periodic displacement are used as the input of the model. The detailed structure of the hybrid model based on LSTM and MDN is shown in Figure. 4, and the hyperparameters such as units in one layer, training epochs, input batch size, and the number of Gaussian distributions are determined by k-fold cross-validation based on the training dataset. The specific operations are as follows: (1) all combinations of hyperparameters were first preferred; (2) then the dataset was randomly split by the k-fold algorithm and the model was trained and validated based on training and validation dataset, respectively. (3) the parameters used in the model with the highest accuracy in the validation dataset are taken as the optimal hyperparameters.
Considering that different parameter initialization affects the final result, this study calculates the model under different parameter initialization conditions 100 times and takes the average values as the final results. This operation is also called the ensemble model prediction, which aims to reduce the model’s prediction uncertainty [14]. Figure 5 shows the mean values of prediction and 90% confidence intervals in the training dataset and testing dataset. It shows that the hybrid model performs well in uncertainty elevation of landslide prediction, i.e., only one point falls outside the 90% confidence interval and the mean values of prediction are close to the ground truth with the root mean square errors (RMSE) of 9.40 and 12.03 mm for training and testing dataset, respectively. Moreover, the proposed model describes a wide confidence interval from 2007.08 to 2008.06, which corresponds to the sharp increase of the landslide displacement. For the Baishuihe landslide, only a small number of data points describe the sharp increase of displacement, so it is reasonable that the prediction uncertainty increases significantly during this period (dashed rectangle in Figure 5), and this phenomenon can not be described by the direct interval estimation model adopted in [7,15], which gave a smaller displacement interval in this period. Figure 6 shows the predictive cumulative displacement along with the ground truth. It should be noted that when the monitoring displacement is greater or less than the confidence intervals, the landslide warning message should be triggered, like the April in 2013 in Figure 6. The results demonstrate that the proposed model is suitable for uncertainty elevation of step-like landslide displacement prediction.
5. Conclusion
This study proposes a hybrid model based on LSTM and MDN to elevate the uncertainty in the landslide displacement prediction. The performance of the model is validated by a typical bank landslide in TGRA, and the results demonstrate that: (i) the prediction uncertainty of landslide prediction can be obtained by introducing Gaussian mixture density to describe the conditional probability of each data point. Compared with the direct prediction of the confidence interval, this model can better describe the greater uncertainty caused by the sharp increase in landslide displacement. (ii) By introducing the LSTM to learn the MDN parameters, the uncertainty evaluation of landslide prediction can be obtained dynamically, which provides more reliable and comprehensive decision-making information for landslide early warning.
6. Acknowledgements
This research has been supported by the China National Key R&D Program during the 13th Five-year Plan Period (grant nos. 2018YFC1505104 and 2017YFC1503103) and the Liao Ning Revitalization Talents Program (grant no. XLYC1807263)

References
[1] Haque U, Paula F, Devoli G, Pilz J, Zhao B, Khaloua A, et al. Science of the Total Environment The human cost of global warming: Deadly landslides and their triggers (1995–2014). Sci Total Environ 2019;673. https://doi.org/10.1016/j.scitotenv.2019.03.415.
[2] Huang R. Large-scale landslides and their sliding mechanisms in China since the 20th century. Yanshilixue Yu Gongcheng Xuebao/Chinese J Rock Mech Eng 2007;26:433–54.
[3] Li H, Xu Q, He Y, Fan X, Li S. Modeling and predicting reservoir landslide displacement with deep belief network and EWMA control charts: a case study in Three Gorges Reservoir. Landslides 2020;17:693–707. https://doi.org/10.1007/s10346-019-01312-6.
[4] Ren F, Wu X, Zhang K, Niu R. Application of wavelet analysis and a particle swarm-optimized support vector machine to predict the displacement of the Shuping landslide in the Three Gorges, China. Environ Earth Sci 2015;73:4791–804. https://doi.org/10.1007/s12665-014-3764-x.
[5] Zhou C, Yin K, Cao Y, Ahmed B. Application of time series analysis and PSO-SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China. Eng Geol 2016;204:108–20. https://doi.org/10.1016/j.enggeo.2016.02.009.
[6] Liu C, Jiang Z, Han X, Zhou W. Slope displacement prediction using sequential intelligent computing algorithms. Meas J Int Meas Confed 2019;134:634–48. https://doi.org/10.1016/j.jmeasurement.2018.10.094.
[7] Wang Y, Tang H, Wen T, Ma J. A hybrid intelligent approach for constructing landslide displacement prediction intervals. Appl Soft Comput J 2019;81:105506. https://doi.org/10.1016/j.asoc.2019.105506.
[8] Lian C, Zeng Z, Member S, Yao W, Tang H, Lung C, et al. Random Hidden Weights. IEEE Trans NEURAL NETWORKS Learn Syst 2016;27:1–13.
[9] Yang B, Yin K, Lacasse S, Liu Z. Time series analysis and long short-term memory neural network to predict landslide displacement. Landslides 2019;16:677–94. https://doi.org/10.1007/s10346-018-01127-x.
[10] Gers FA, Schmidhuber J. Recurrent nets that time and count. Proc. Int. Jt. Conf. Neural Networks, vol. 3, 2000, p. 189–94. https://doi.org/10.1109/ijcnn.2000.861302.
[11] Li D, Yin K, Leo C. Analysis of Baishuihe landslide influenced by the effects of reservoir water and rainfall. Environ Earth Sci 2010;60:677–87. https://doi.org/10.1007/s12665-009-0206-2.
[12] Miao F, Wu Y, Xie Y, Li Y. Prediction of landslide displacement with step-like behavior based on multialgorithm optimization and a support vector regression model. Landslides 2018;15:475–88. https://doi.org/10.1007/s10346-017-0883-y.
[13] Zhang L, Shi B, Zhu H, Yu XB, Han H, Fan X. PSO-SVM-based deep displacement prediction of Majiagou landslide considering the deformation hysteresis effect. Landslides 2020. https://doi.org/10.1007/s10346-020-01426-2.
[14] Cheng H, Chen H, Li Z, Cheng X. Ensemble 1-D CNN diagnosis model for VRF system refrigerant charge faults under heating condition. Energy Build 2020;224:110256. https://doi.org/10.1016/j.enbuild.2020.110256.
[15] Xing Y, Yue J, Chen C. Interval Estimation of Landslide Displacement Prediction Based on Time Series Decomposition and Long Short-Term Memory Network. IEEE Access 2020;8:3187–96. https://doi.org/10.1109/ACCESS.2019.2961295.