Research on Highway Slope Monitoring Data Prediction Based on Long Short-term Memory Network

Jianbo Zhang¹, Ting Peng²*, Linhai Mo¹, Xu Hu³ and Xiaoling WANG⁴

¹School of Highway, Chang'an University, Xi'an, Shaanxi, 710064, China
²Special Region Highway Engineering Lab, Chang'an University, Xi'an, Shaanxi, 710064, China
³China Design Group Co., Ltd., Nanjing, Jiangsu, 210014, China
⁴Xi'an Branch of the People's Bank of China, Xi'an, Shaanxi, 710064, China

*Corresponding author’s e-mail: pt@chd.edu.cn

Abstract. This paper first analyzes the basic structure and operation principle of Long Short-Term Memory Network (LSTM), and combines the characteristics of road slope monitoring data to determine the data form of the input layer and the type of activation function. Then, combined with the slope displacement monitoring data, a slope monitoring data prediction model based on LSTM is constructed, and the main structural parameters of the LSTM are optimized to predict the slope monitoring data. Finally, the data prediction results are analyzed, and the reliability of the prediction results is verified by analyzing the influence of the number of monitoring data samples on the prediction results.

1. Introduction

By analyzing the monitoring results of highway slope disasters⁸, it can be found that the monitoring data sequences of slope monitoring points are mostly time series with monotonous increasing or monotonous decreasing with time, which has certain regularity. Therefore, this paper extracts information on slope deformation by monitoring data. However, the data sequence obtained by the monitoring can only indicate the deformation history and variation law of the research object in the past time period, and can’t predict the future deformation trend of the slope body⁹. In order to solve this problem, this paper uses Long Short-Term Memory Network (LSTM) to construct the LSTM prediction model for slope monitoring data from real-time monitoring data, and obtain the predicted value of monitoring data in a certain period of time in the future, through monitoring data and forecasting data. Combined with organic monitoring data, a new monitoring data series is formed to evaluate the safety of highway slope monitoring points.

2. Long Short-Term Memory Network (LSTM)

The basic structure of the LSTM is shown in Figure 1. The "gate" used in LSTM⁸ is actually a fully connected layer. By inputting a set of vectors, the real map vector in the corresponding interval is output. In this paper, the forget gate, the input gate, the output gate and the weight matrix of the cell state are represented by W_f, W_i, W_o and W_c, and the bias terms are represented by b_f, b_i, b_o, and b_c. The combined vector [h_{t-1}, x_t] includes both the implicit state of the previous moment and the input value of the current moment. The two sets of vectors propagate in the same way during forward
propagation, but there are formulas for each of the backward propagation. Therefore, each weight matrix can be split into two matrices, denoted by \( W_{Fh} \) and \( W_{Fx} \), \( W_{Ih} \) and \( W_{Ix} \), \( W_{Oh} \) and \( W_{Ox} \), \( W_{Ch} \) and \( W_{Cx} \)[8].

**Figure 1. Schematic diagram of the basic structure of the LSTM.**

3. Prediction of slope monitoring data based on LSTM

This paper trains and tests various types of monitoring data collected by the monitoring system to analyze the impact of four important parameters of learning rate, batch, epoch and neuron of hidden layers on the LSTM. The slope monitoring data prediction process based on the LSTM is shown in Figure 2.

**Figure 2. Slope monitoring data prediction flow chart based on LSTM.**

The main research processes involved in this paper are all done under the TensorFlow framework. The computer parameters are Intel(R) Core(TM) i7-7700HQ CPU, 16.0GB RAM, GeForce GTX 1050, 2.0GB GPU.
In this paper, a cross-validation method is used to select 7100 sets of slope displacement monitoring data for predictive model construction. Among them, 4734 sets are randomly selected as training samples, 2366 sets of data are used as test samples, and the ratio of the number of train set to the number of test set is about 2:1, and the train set and test set are not duplicated to ensure that the test set is valid.

4. Parameter Analysis of LSTM

This section specifically studies the effects of structural parameters such as learning rate, batch, epoch, and neuron of hidden layers on the LSTM. In order to verify the training effect of the neural network, the mean absolute error (MAE) of each iteration is used when different parameters are used, and the optimal parameters are determined by comparing the output value of the MAE and the convergence speed.

4.1. Learning rate

The default setting of the parameters in this paper is 0.35 for drop-out rate, 32 for batch, 500 for epoch, 3 layers for hidden layer, and 15 for neuron of hidden layers. Firstly, the learning rate of five different magnitudes of 0.0001, 0.001, 0.01, 0.1, and 1 are selected to obtain the MAE, the test loss and the train loss. As shown in Table 1.

| Learning rate | 0.0001 | 0.001 | 0.01 | 0.1 | 1 |
|---------------|--------|-------|------|-----|---|
| MAE (Train set) | 0.261 | 0.161 | 0.160 | 0.160 | 0.491 |
| Train loss | 0.112 | 0.045 | 0.045 | 0.045 | 0.376 |
| MAE (Test set) | 0.217 | 0.129 | 0.128 | 0.137 | 0.437 |
| Test loss | 0.105 | 0.035 | 0.033 | 0.036 | 0.329 |

The training result is relatively optimal when the learning rate is 0.01, and the effect is second when the learning rate is 0.001. When the learning rate is 0.0001 and 0.1, the training result is between the best and the worst. When the learning rate is 1, the prediction model falls into the local optimal solution.

Based on the above analysis results, the 0.001 test set and the 0.050 test set with better MAE were selected as the data interval for secondary selection. In the step of 0.005, the learning rate of the LSTM is updated in the interval by successive incremental values. 0.001, 0.005, 0.010, 0.015, 0.020, 0.025, 0.030, 0.035, 0.040, 0.045, and 0.050 are selected as the learning rate.

| Learning rate | 0.001 | 0.005 | 0.010 | 0.015 | 0.020 | 0.025 | 0.030 | 0.035 | 0.040 | 0.045 | 0.050 |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Epoch / times | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 |
| Total iterative number / thousand times | 103.5 | 103.5 | 103.5 | 103.5 | 103.5 | 103.5 | 103.5 | 103.5 | 103.5 | 103.5 | 103.5 |
| Running time / min | 481 | 446 | 425 | 421 | 431 | 418 | 419 | 400 | 411 | 425 | 443 |
| MAE | 0.129 | 0.128 | 0.128 | 0.128 | 0.127 | 0.127 | 0.126 | 0.126 | 0.124 | 0.128 | 0.132 |
| Iterative to MAE stable / times | 251 | 212 | 197 | 172 | 145 | 133 | 128 | 115 | 112 | 124 | 133 |
Based on Table 2, the size of the learning rate does not affect the epoch or the total iterative round. The two are independent of each other and do not affect each other. The running time of 11 set of learning rate is listed in the table. When the learning time is 0.035, the running time is the shortest, and when the learning rate is 0.001, the running time is the longest. In general, the change of the learning rate will directly affect the running time. The running time increment is more obvious as the learning rate is too small. When the learning rate is 0.001, the error converges to stability when iterating 251 times, and when the learning rate is 0.040, iteration can be converged until 112 times. Compared to the learning rate of 0.040, its running time and MAE are at a disadvantage.

Figure 3 is the fluctuation curve of the loss when the learning rate is 0.040. In the figure, the train loss and the test loss change trend are basically the same, and finally reach a relatively stable state, indicating that the input monitoring data sequence is valid and the model has converged. Considering Figure 3 and Table 2, when the learning rate is 0.040, the train set and test set are more stable, and the running time and convergence speed are relatively short. Therefore, the learning rate is 0.040.

![Figure 3. The fluctuation curve of the loss when the learning rate is 0.040.](image)

### 4.2. Batch

In order to study the influence of the batch on the predicted value of the monitoring data, the learning rate is 0.01, and the other default parameters remain unchanged. The value is incremented in the interval of 16 steps, and 16, 32, 48, 64, 80, 96, 112, 128, 144 and 160 are used as batch to analyze the influence of the batch. The training results are shown in Table 3 below.

The size of the batch directly affects the total iterative number, which is inversely proportional. When the epoch is kept 500 times, the total iterative number decreases rapidly as the batch increases. The running time of 10 sets of batch is listed in the table, and the running time decreases as the batch increases. The increase in batch can significantly reduce the running time and improve the operating efficiency. The increase in the batch causes the number of iterations required for error convergence to increase gradually, while the MAE of the output decreases, but the trend is not obvious. When the batch is taken as 160, the MAE of the output is the smallest, which is 0.120, and the error reaches a steady state from the iteration to the 212 times.

| Parameter               | Batch | 16 | 32 | 48 | 64 | 80 | 96 | 112 | 128 | 144 | 160 |
|-------------------------|-------|----|----|----|----|----|----|-----|-----|-----|-----|
| Epoch / times           |       | 500| 500| 500| 500| 500| 500| 500  | 500  | 500  | 500  |
| Total iterative number / thousand times |       | 207.5| 103.5| 69.0| 51.5| 41.5| 34.5| 29.5 | 25.5 | 23.0 | 20.5 |
| Running time / min      |       | 607 | 559| 511| 459| 426| 391| 363  | 325  | 300  | 282  |
| MAE                     |       | 0.133| 0.129| 0.131| 0.129| 0.126| 0.126| 0.124 | 0.122 | 0.123 | 0.120 |
Iterative to MAE stable/
times 10^3

| Learning rate | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 |

The above rules are basically consistent with the existing theory. Research shows that as the batch becomes larger, the data volume of each group becomes more and the correct rate of the model becomes higher, but the excessive value of the batch reduces the update speed of the parameters, and the randomness of the gradient update decreases.

Considering Table 3, the MAE of the output is the smallest when the batch is 160. Considering that the smaller the batch, the longer the program runs, so the value of the batch is 160. Figure 4 is the loss fluctuation curve when the batch is 160. The train loss and the test loss change trend are basically the same, and finally reach a relatively stable state, indicating that the input monitoring data sequence is valid and the model has converged.

4.3. Epoch

One iteration in this section is set to completely traverse the data set once. In order to study the influence of the epoch on the predicted value of the monitoring data, the learning rate is 0.040, the batch is 160, the other default parameters remain unchanged. The value is incremented in the interval of 100 steps, and the epoch is taken as 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000. The training results are shown in Table 4 below.

(1) The size of the epoch also has an effect on the total iterative number, which is directly proportional. When the batch remains 160, the total iterative number increases as the epoch increases.

(2) The running time of 10 different epoch is listed in the table, and the running time is significantly extended with the increase of the epoch. During the training, it was observed that even if the same set of parameters were used for the repeated test, the running time would still have a time deviation of several seconds to several minutes. One of the reasons for this phenomenon is the influence of the drop-out setting on the model construction. It also may be Computer hardware causes.

(3) The increase in the epoch has a great influence on the error convergence result. With the current train set and test set, the MAE can’t converge when the epoch is 100, 200, 900, and 1000. When the epoch is 100, the MAE shows a downward trend, but the number of iterations is too small, and finally it cannot converge. When the epoch is 900, the LSTM falls into a local optimal solution, and the predicted monitoring value is greatly deviated from the actual monitored value. The error of the predicted value can’t converge, and the prediction result is not credible. When the epoch is 1000, the error of the test is gradually reduced in the early stage, and there is a large fluctuation in the later stage, which is difficult to stabilize.

The MAE of the other 7 experimental groups was stable. When the epoch is 400, the MAE is the smallest, which is 0.121. When the epoch is 300, the minimum number of algebras required for iterative to MAE stability is 207 times. When the epoch is 600 and 700, the error curves of the two are...
quite different. It may be the random selection of neurons during training produces different operation results.

| Parameter          | Epoch 100 | Epoch 200 | Epoch 300 | Epoch 400 | Epoch 500 | Epoch 600 | Epoch 700 | Epoch 800 | Epoch 900 | Epoch 1000 |
|--------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
| Batch              | 160       | 160       | 160       | 160       | 160       | 160       | 160       | 160       | 160       | 160        |
| Total iterative number / times | 4100      | 8200      | 12300     | 16400     | 20500     | 24600     | 28700     | 32800     | 36900     | 41000      |
| Running time / min | 52        | 112       | 174       | 227       | 285       | 347       | 396       | 455       | 513       | 577        |
| MAE                | 0.246     | 0.137     | 0.126     | 0.121     | 0.123     | 0.127     | 0.137     | 0.136     | 0.216     | 0.194      |
| Iterative to MAE stable / times | ----      | ----      | 207       | 209       | 212       | 217       | 215       | 221       | ----      | ----       |
| Learning rate      | 0.040     | 0.040     | 0.040     | 0.040     | 0.040     | 0.040     | 0.040     | 0.040     | 0.040     | 0.040      |

Based on Table 4, the MAE is the smallest when the epoch is 400, and the test result is better. The running time is shorter when the epoch is 300. Taking into account the accuracy priority principle, the epoch takes 400. Figure 5 is the fluctuation curve of the loss when the epoch is 400. In the figure, the train loss and the test loss change trend are basically the same, and finally reach a relatively stable state, which indicating that the input monitoring data sequence is valid and the model has converged.

4.4. Neuron of hidden layers

In order to study the influence of the neuron of hidden layers on the predicted value of the monitoring data, the learning rate is 0.040, the batch is 160, and the epoch is 400. The other default parameters are kept unchanged, and the value is incremented in the interval [1, 50] in steps of 5.

Table 5 lists the training states of the different neuron of hidden layers. Analysis Table 5 can find that when the neuron of hidden layers is too small, it is difficult for the LSTM to obtain feature information from the monitoring data sequence. It is impossible to mine the variation law of the existing data, and the performance is poor. When the neuron of hidden layers is too large, the systematic error of the network may be reduced. On the one hand, the running time of the network is prolonged. On the other hand, the training is easy to fall into the local optimal solution and the overall optimal result is not obtained. Therefore, when selecting the neuron of hidden layers, the accuracy requirements of data prediction and the complexity of the LSTM should be fully considered, and the total number of samples of the data sequence should be combined to determine the effective neuron of hidden layers.
### Table 5. Training states of different neuron of hidden layers.

| Neuron of hidden layers | Parameter          | Value       | Value   | Value   | Value   | Value   | Value   | Value   | Value   | Value   |
|-------------------------|--------------------|-------------|---------|---------|---------|---------|---------|---------|---------|---------|
|                         | Running time / min | 194         | 206     | 218     | 227     | 235     | 249     | 261     | 276     | 283     | 294     | 311     |
|                         | MAE                | 0.806       | 0.185   | 0.139   | 0.121   | 0.121   | 0.124   | 0.122   | 0.137   | 0.243   | 0.261   | 0.227   |
|                         | Iterative to MAE stable/ times | ---- | ---- | 183 | 209 | 217 | 229 | 241 | 265 | ---- | ---- | ---- |
|                         | Batch              | 160         | 160     | 160     | 160     | 160     | 160     | 160     | 160     | 160     | 160     |
|                         | Epoch / times      | 400         | 400     | 400     | 400     | 400     | 400     | 400     | 400     | 400     | 400     |
|                         | Total iterative number/ times | 16400 | 16400 | 16400 | 16400 | 16400 | 16400 | 16400 | 16400 | 16400 | 16400 |
|                         | Learning rate      | 0.040       | 0.040   | 0.040   | 0.040   | 0.040   | 0.040   | 0.040   | 0.040   | 0.040   | 0.040   |

At present, there is no unified and reliable theory for determining the neuron of hidden layers. Generally, the same set of data is repeatedly tested by trial and error method. Compare the running results of different parameters and comprehensively select the most suitable parameter values.

From the table 5, when the neuron of hidden layers is 15 or 20, the MAE is the smallest. Considering the relationship between the neuron of hidden layers and the running time, convergence speed, the neuron of hidden layers is determined to be 15. Figure 6 is the fluctuation curve of the loss when the neuron of hidden layers is 15. In the figure, the train loss and the test loss change trend are basically the same, and finally reach a relatively stable state, indicating that the input monitoring data sequence is valid and the model has converged.

![Figure 6. Loss fluctuation curve when the neuron of hidden layers is 15.](image)

### 5. Analysis of prediction results

Based on the parameter analysis results of the LSTM from 4.1 to 4.4, the final learning rate is 0.040, the batch is 160, the epoch is 400, the hidden layer has three layers, and the neuron of hidden layers is 15, the drop-out rate is 0.35 as the default parameter of the LSTM. Run the neural network again. The running result is shown in Figure 7. The comparison curve between the predicted displacement value and the actual monitored displacement value is plotted. At this time, the minimum MAE of the system is 0.120.
Figure 7. Comparison of LSTM predicted displacement values and actual monitored displacement values.

6. Conclusion
In this paper, a slope monitoring data prediction model based on the Long Short-Term Memory Network (LSTM) is constructed, and four structural parameters, such as learning rate, unit value, iteration cycle and number of hidden layer nodes, are optimized in combination with 7,100 groups of slope displacement monitoring data. The optimization results show that when the selection learning rate is 0.040, the unit value is 160, and the iteration rounds are 400, there are three hidden layers, the number of hidden layers is 15, and the drop-out rate is 0.35 as the optimal parameter combination, the LSTM model has the best prediction effect on the current monitored displacement data, and the minimum average absolute error is reduced to 0.120. Among the four parameters, the learning rate and the number of hidden layer nodes have the greatest influence on the LSTM model, while the influence of cell value and iteration rounds is relatively small.

References
[1] Wang, W.D., Liu, C., Li, D.H. (2015) Gis-based Highway Slope Risk Analysis and Early Warning System Research [J]. Journal of Central China Normal University (Natural Science Edition), 49(03): 452-459.
[2] Yan, E., Song, K., Li, H. (2010) Applicability of time domain reflectometry for Yuhuangge landslide monitoring[J]. Journal of Earth Science, 21(6): 856-860.
[3] Ohnishi, Y., Nishiyama, S., Yano, T. (2006) A study of the application of digital photogrammetry to slope monitoring systems[J]. International Journal of Rock Mechanics and Mining Sciences, 43(5): 756-766.
[4] Jiang, G., Lin, L.S., Liu, Z.D., et al. (2000) Grey prediction model of slope Deformation [J]. Rock and Soil Mechanics, 21(3): 244-246,251.
[5] Yan, H.B., Yang, Q., Ren, C., et al. (2017) Prediction of SLOPE based on BP Neural Network based on EEMD [J]. Hydraulic Power Generation, 43(7): 37-40.
[6] Zhang, Z.H., Yuan, M.K., Deng, J.H., et al. (2014) Slope displacement prediction based on improved Gray-time series analysis time-variation model [J]. Journal of Rock Mechanics and Engineering, 33(S2): 3791-3797.
[7] Yang,B.B., Yin, Q.L., Du, J. (2018) Dynamic prediction model of landslide displacement based on time series and long-term memory network [J]. Journal of Rock Mechanics and Engineering, 37(10): 2334-2343.
[8] Graves, A. (2012) Long Short-Term Memory[J].
[9] Chen,W.Y., Liao, Y.F., Chen, S.H. (1995) Speech recognition with hierarchical recurrent neural networks[J]. Pattern Recognition, 28(6):795-805.
[10] Xu, S., Niu, R. (2018) Displacement prediction of Baijiabao landslide based on empirical mode decomposition and long short-term memory neural network in Three Gorges area, China[J]. Computers & Geosciences, 111:87-96.