Ma, Enlin; Janiszewski, Mateusz; Torkan, Masoud

Deep learning methods for underground deformation time-series prediction

Published in:
Expanding Underground - Knowledge and Passion to Make a Positive Impact on the World- Proceedings of the ITA-AITES World Tunnel Congress, WTC 2023

DOI:
10.1201/9781003348030-334

Published: 12/04/2023

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Ma, E., Janiszewski, M., & Torkan, M. (2023). Deep learning methods for underground deformation time-series prediction. In G. Anagnostou, A. Benardos, & V. P. Marinos (Eds.), *Expanding Underground - Knowledge and Passion to Make a Positive Impact on the World- Proceedings of the ITA-AITES World Tunnel Congress, WTC 2023: Proceedings of the ITA-AITES World Tunnel Congress 2023 (WTC 2023), 12-18 May 2023, Athens, Greece* (1st Edition ed., pp. 2775-2781). CRC Press. https://doi.org/10.1201/9781003348030-334

Expanding Underground. Knowledge and Passion to Make a Positive Impact on the World – Anagnostou, Benardos & Marinos (Eds)

Deep learning methods for underground deformation time-series prediction

E. Ma
School of Highway, Chang'an University, Xi'an Shaanxi province, China

M. Janiszewski & M. Torkan
Department of Civil Engineering, School of Engineering, Aalto University, Espoo Helsinki, Finland

ABSTRACT: Prediction is a vague concept that is why we need to conceptualize it specifically for underground deformation time-series data. For this impending issue, this paper employs an advanced deep learning model Bi-LSTM-AM to address it. The results show its applicability for practical engineering. The proposed model is compared with other basic deep learning models including long short-term memory (LSTM), Bi-LSTM, gated recurrent units (GRU), and temporal convolutional networks (TCN). These models cover the most common three forms of deep learning for time-series prediction: recurrent neural networks (RNN) and convolutional neural networks (CNN). This research is supposed to benefit the underground deformation time-series prediction.

Keywords: underground engineering, time-series, deep learning, deformation prediction, machine learning

1 INTRODUCTION

In underground engineering, structural deformation is an intuitive, easy to obtain, and can effectively reverse the structural health status and surrounding rock and soil. Here, underground deformation can refer to the deformation of underground structure or the underground constructing-induced deformation of surface structures. Predicting deformation is to predict the safety state of the structure in the future and prevent potential risks. Many methods can be used to predict structural deformation. The expected deformation of the structure can be calculated through analytical formulas and numerical simulation. However, due to the complexity of geotechnical materials, it is difficult to achieve high accuracy only by mechanical means. another method to predict the deformation is according to the deformation monitoring data. The monitoring data is usually in the form of time series, therefore the time-series method can be used to predict the deformation. The prediction results can also be updated with the accumulation of monitoring data. The traditional time-series method is ARMA model (*Fan & Jian* 2019). Recently, the development of machine learning has brought new approaches to deformation prediction.

For underground structure deformation, there are two commonly used machine learning prediction models: (1) classical models: more powerful than mathematical statistics with enough data, including support vector regression (SVR), extreme learning machines (ELM), XGBoost, etc (*Ahmed et al.* 2020). (2) Deep learning models: it refers to deep neural networks or, in the broad sense, any iterative multi-layer machine learning models. It has strong generalization ability, mainly including two types of methods: (a) recurrent neural network (RNN):

DOI: 10.1201/9781003348030-334

long short-term memory (LSTM), gated current unit (GRU), DeepAR, etc. (*Wang et al.* 2021; *Yu et al.* 2019), and (b) convolutional neural network (CNN): temporal convolutional network (TCN), WaveNet, etc (*Livieris et al.* 2020). These methods provide more possibilities for the deformation prediction of underground structures. Further research is focused on the model optimization, which is mainly divided into: (1) combining LSTM, attention mechanism (AM) and other methods to optimize the model structure (*Vaswani et al.* 2017); (2) Introducing particle swarm optimization, drosophila, gray wolf and other algorithms to improve the model parameters.

In short, this research uses deep learning models to underground time-series deformation, which can be more practical for understanding the temporal development of deformation during construction and operation. The remainder of this paper is organized as follows. Section 2 conceptualizes the time-series prediction. Section 3 employs a deep learning model Bi-LSTM-AM to predict the time-series deformation of a underground project. Section 4 shows the results, which are compared with other basic deep learning models including LSTM, Bi-LSTM, GRU, and TCN. Section 5 draws up the conclusions.

2 CONCEPTUALIZATION

In order to distinguish from other prediction methods, the process of predicting underground construction deformation by time series method is conceptualized via its mathematical form. For the time-series deformation data collected by each sensor, the prediction task is to obtain the deformation in a certain future period through the existing monitoring data:

$$D_{t+1:t+q} = F\left[x_{(x-p+1):t}\right] \tag{1}$$

where $D_{t+1:t+q}$ is predicted deformation from time t+1 to t+q; F is the mapping relation; and $x_{(x-p+1):t}$ are monitoring data from time $x - p + 1$ to t.

According to this definition, q is the prediction step, that is, the deformation after time q is predicted; P is the observation length, that is, p monitoring data from time $t - p + 1$ to time t are selected for prediction.

In order to obtain above mapping relationship, model training is required to transform the time series vector of monitoring data $x_{1:t} = \{x_1, x_2, \ldots, x_t\}$ into input matrix $\boldsymbol{A}$ and output matrix $\boldsymbol{B}$:

$$A = \begin{bmatrix} x_1 & x_2 & \cdots & x_p \\ x_2 & x_3 & \cdots & x_{p+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{t-q-p+1} & x_{t-q-p+2} & \cdots & x_{t-q} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{t-q-p+1} \end{bmatrix} \tag{2}$$

$$B = \begin{bmatrix} x_{p+1} & x_{p+2} & \cdots & x_{p+p} \\ x_{p+2} & x_{p+3} & \cdots & x_{p+1+q} \\ \vdots & \vdots & \vdots & \vdots \\ x_{t-q+1} & x_{t-q+2} & \cdots & x_t \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{t-q-p+1} \end{bmatrix} \tag{3}$$

The training of prediction model refers to solving a mapping relationship $\boldsymbol{F}$ to make c and $\boldsymbol{b}$ closest to the whole, that is, the overall error is the minimum. The model training error ε is:

$$\varepsilon = \sqrt{\sum_{i=1}^{n} \frac{\left[\boldsymbol{F}(a_i) - \boldsymbol{b}_i\right]^2}{n}} \tag{4}$$

where n = t-q-p+1.

The above process is the fundamental part of time-series prediction. More complicated prediction can be conducted through considering more factors. First, while predicting deformation, more time-series data can be incorporated such as temperature and water pressure. In this condition, the matrix $\boldsymbol{B}$ is still the same but in matrix $\boldsymbol{A}$, each x_i should be a array or vector that represents more information than deformation. Second, some static factors which do not change with time can be considered such as the geometry information of the monitoring section.

For the time series prediction problem, all kinds of machine learning algorithms are solving the above mapping relationship $\boldsymbol{F}$, and this standard form is the essential feature of time-series prediction with machine learning which is different from other methods.

3 MODEL ESTABLISHMENT

3.1 *Data pre-processing*

The deformation time-series was derived from an open-access data set. The deformation inside a rock was monitored, and the temperature was recorded meanwhile. For each monitoring place, a borehole extensometer was anchored in the rock mass at the end of the borehole at a depth of 8 m from the cliff wall. Three displacement-measuring points at depths of 2 m, 4 m and 6 m from the cliff wall and temperature measuring points at 2 m and 6 m depth.

In this research, the displacement at depth of 6 m in the rock mass is chosen as the prediction target. The input time series include the displacements at depths of 2 m, 4 m and 6 m and temperature at 6 m depth. Therefore, the input dimensions are 4. In the original data set, the displacement and temperature are monitored once per hour (from 22.07.2010 to 08.04.2015) but few records are missing. The prediction task is to use the data (the displacements at depths of 2 m, 4 m and 6 m and temperature at 6 m depth) of the past three days to predict the displacement at 6 m depth of the next day.

According to Equations (2) and (3), the original data set is transformed into input array and output array. After removing samples that contain missing data, the shape of input array is (35169, 72, 4). 35169 is the sample size; 72 is the length of time series; and the 4 is the input dimensions or numbers of the time series. The output array shape is (35169, 24), and 24 is the length of time series, i.e., the displacements of the next 24 hours.

In a machine learning task, there should be training data, validation data and test data. Training data is used for training parameters of model. Test data is used to test the generalization performance of the final model that has been trained. It should be noted that before the test, the model parameters have been determined, and these parameters do not change after the test. Validation data is used to check the performance of the model (this is the same as the test set), but the model parameters, mainly the hyper-parameters, can be adjusted in turn according to the inspection results. Since the sample size is very large in this case, only 5% of the samples are used for training, and 1% for validation, and 94% for test. Before dividing samples into training, validation and test data, the samples are disordered randomly to make each data set has sufficient time span.

The last step before training is to normalize data set:

$$\widehat{x} = \frac{x - \overline{x}}{\sigma} \tag{5}$$

where $\widehat{x}$ is the data after normalization; $\overline{x}$ is the mean of x; σ is the standard deviation of x. It should be not that the validation and test data cannot be involved when calculating $\overline{x}$ and σ, because they are unknown when we train a model in practical engineering.

3.2 *Experimental setup and parameter configuration*

We proposed a Bi-LSTM-AM model to predict. In short, Bi-LSTM is used for processing time series data, and AM is to estimate the output weight of Bi-LSTM at each time step. The principle of this model can be briefly described in Figure 1. The networks are developed in Keras. The Keras framework is encapsulated well but its customization is limited. In order to realize the attention mechanism, its neural network layer is used as a function call, and the object-oriented programming model is adopted. After grid search optimization algorithm, the model structure is determined and shown in Table 1.

The batch size is 128; epoch number is 100; optimizer is Adam; loss metric is mean square error. The training loss and validation loss along with epochs are shown in Figure 2. The training loss and validation loss both decrease steadily, and do not fluctuate apparently after a few epochs. After 100 epochs, the model is obtained and will be tested by test data set.

4 RESULTS AND DISCUSSION

4.1 *Results of Bi-LSTM-AM*

The model was used to test 33059 size of data set. For each sample, we can have a time series with length of 24, which means the predicted following 24-hour displacements. The mean absolute error (MAE) and symmetric mean absolute percentage error (SMAPE) are chosen to evaluate the model performance, as shown in Figure 3. As we can see, the total errors of both MAE and SMAPE are very low, however, it cannot indicate that the prediction performance must be good.

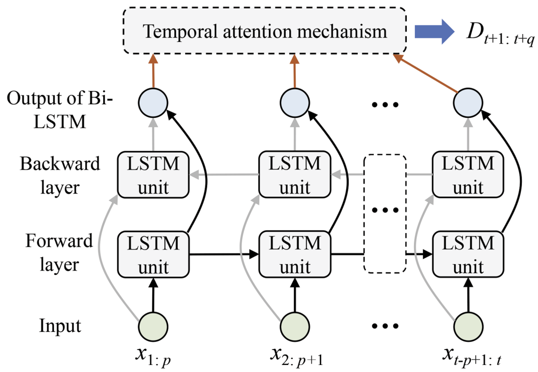

Figure 1. The principal and process of Bi-LSTM-AM model with one layer of Bi-LSTM.

Table 1. The model structure of Bi-LSTM-AM.

Layer*	Output shape	Connected to
Input layer	(None, 72, 4)	
Bi-LSTM layer 1	(None, 72, 256)	Inpur layer
Bi-LSTM layer 2	(None, 72, 512)	Bi-LSTM layer 1
Bi-LSTM layer 3	(None, 72, 128)	Bi-LSTM layer 2
Permute 1	(None, 128, 72)	Bi-LSTM layer 3
Dense layer 1	(None, 128, 72)	Permute 1
Permute 2 (Attention vector)	(None, 72, 128)	Dense layer
Attention multiply	(None, 72, 128)	Bi-LSTM layer 3 and Permute 2 (Attention vector)
Flatten layer	(None, 9216)	Attention multiply
Dropout layer	(None, 9216)	Flatten layer
Dense layer 2	(None, 128)	Dropout layer
Dense layer 3	(None, 64)	Dense layer 2
Dense layer 4	(None, 24)	Dense layer 3

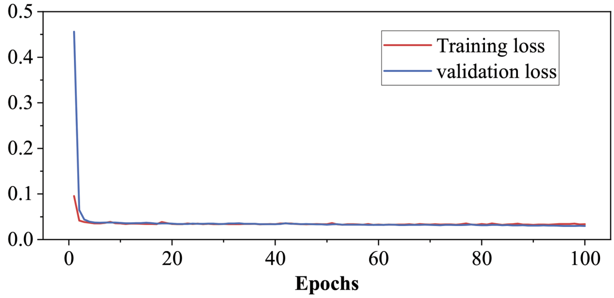

Figure 2. The training and validation loss during model training.

This is because in this case, the displacement time series do not change fiercely in short term. Currently, there is not common metric to reasonably evaluate the model performance suitable for all conditions with different features that how much time series change in concerned time steps. A more reasonable evaluation indicator should be proposed, which is the task in our further studies. But overall, the errors are low and acceptable in practical underground engineering.

Another perspective to evaluate the model performance is the change of errors on different prediction time steps. In the past, one problem when using ARIMA or classical machine learning methods to predict time series is that the error will increase rapidly with the increase of prediction time step. This is caused by the cumulative error. However, in Figure 3, only the errors

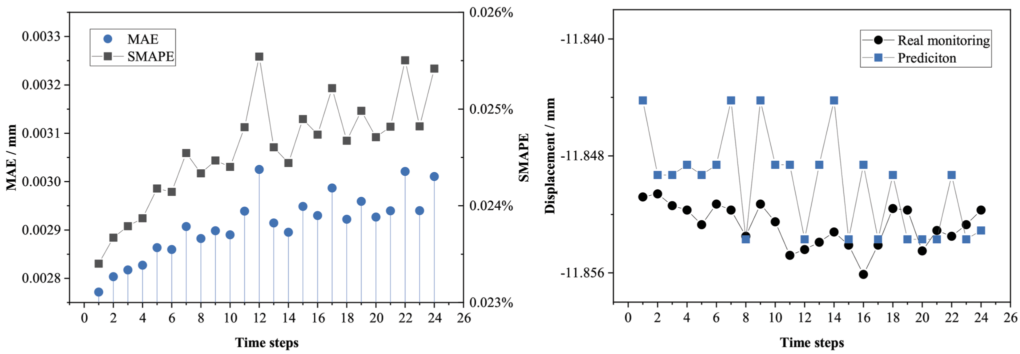

Figure 3. The MAE and SMAPE of test results on 24 prediction time steps for all samples, and a 24 time-step prediction results of one sample.

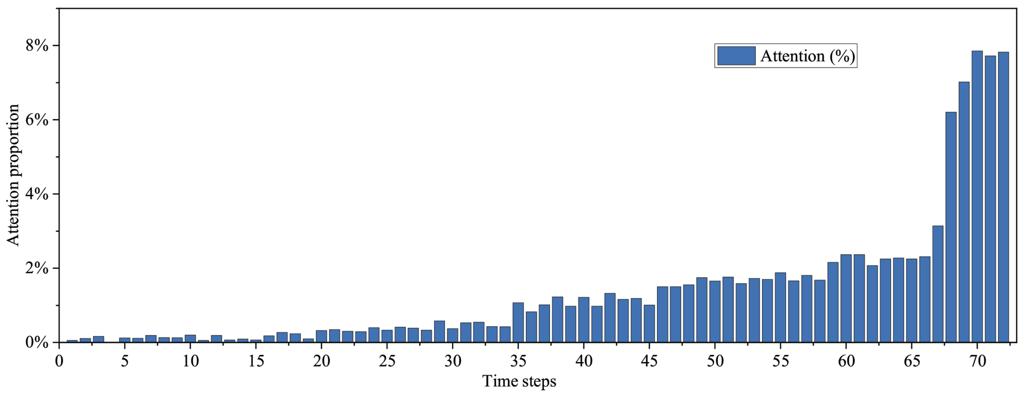

Figure 4. The weights of each time step in training data determined by AM.

on the first 10 prediction time steps are low and keeping increasing, while on the rest 14 time steps (from 11 to 24), the errors do not have a increase trend anymore. It reveals the potential of deep learning model to predict on long time steps without an obvious non-convergence.

The temporal attention mechanism is used in this research. It can calculate the weights of the output of "Bi-LSTM layer 3" on each time step. As shown in Figure 4, when predicting the next 24-hour displacements using 72-hour data, the last 5 time steps from 68 to 72 has the most weights. AM can make the whole model focus on the crucial time steps, which will have more influence on prediction results. In big data training task, AM is important to increase the efficiency and accuracy of the model.

4.2 *Discussion*

The MSE, MAE and SMAPE of Bi-LSTM-AM are compared with other basic deep learning models, as shown in Figures 5 and 6. First, Bi-LSTM-AM model achieves a highest accuracy on the test data set. But all RNN models including Bi-LSTM-AM, LSTM, Bi-LSTM, and GRU have similar performance. TCN, which belongs to a type of 1D CNN models, does not have a advantage in this research. In addition, the Bi-LSTM-AM model has more obvious advantage when the prediction steps are large.

5 CONCLUSION

This research proposed a Bi-LSTM-AM model to predict a underground time-series displacement. The prediction accuracy is acceptable. The attention mechanism has increased

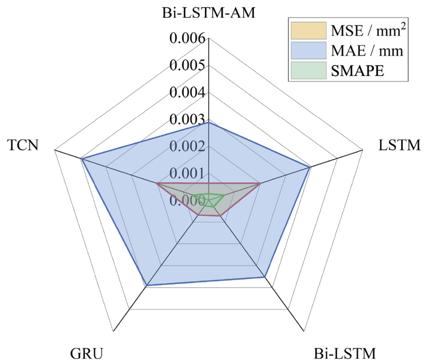

Figure 5. The MSE, MAE and SMAPE of each deep learning models (from prediction time step 1 to 12).

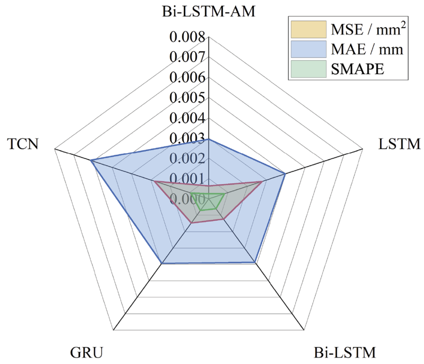

Figure 6. The MSE, MAE and SMAPE of each deep learning models (from prediction time step 13 to 24).

the performance of prediction, compared with other basic deep learning models. The main conclusions are as follows:

1) The deep learning models have strong applicability. This technique allows data set with different values and features to be trained and subsequently predicted accurately.
2) The results of Bi-LSTM-AM are with a prediction error MAE around 0.003 mm, which is acceptable considering the value variation range.
3) The introduction of attention mechanics enables the model to calculate a large window-scale and prediction-step problem. Although on the whole, the effect of AM on prediction accuracy is limited, but the effect is more significant in the farther prediction steps.

REFERENCES

Ahmed, N.K., Atiya, A.F., Gayar, N.E. & El-Shishiny, H., 2010. An empirical comparison of machine learning models for time series forecasting. *Econometric reviews* 29(5-6):594–621.

Fan, F. & Jian, L.I., 2019. A Prediction Model of Tunnel Surrounding Rock Deformation Based on Induced Ordered Weighted Averaging Theory and its Application. *In IOP Conference Series: Earth and Environmental Science* 267(4):042098). IOP Publishing.

Livieris, I.E., Pintelas, E. & Pintelas, P., 2020. A CNN–LSTM model for gold price time-series forecasting. *Neural computing and applications* 32(23):17351–17360.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. & Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems* 30.

Wang, R., Li, D., Chen, E.J. & Liu, Y., 2021. Dynamic prediction of mechanized shield tunneling performance. *Automation in Construction* 132:103958.

Yu, Y., Si, X., Hu, C. & Zhang, J., 2019. A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation* 31(7):1235–1270.