Time series analysis and LSTM-LighGBM Model for groundwater forecasting

Chongzhi Wu*

PhD student, School of Civil Engineering, Chongqing University, Chongqing 400045, China. E-mail: iwuchongzhi@cqu.edu.cn

Songlin Liu*

PhD student, School of Civil Engineering, Chongqing University, Chongqing 400045, China. E-mail: songlinl@usc.edu

*** The authors contributed equally to this work**

Environmental concerns are always an important issue during construction periods. The detailed comparison analysis of actual measurement about environmental parameters under construction versus that under natural situation is vital for carrying about the correct actions to mitigate the negative effects of constructing activities on nature and buildings along the project. Generally, this is an impossible mission as there is no way to obtain the unaffected values once the project starts, but the birth of new technology artificial intelligence (AI) finally provides the solution.

In order to analyze and reduce the negative impacts of the planned 40km long megaproject on nature, a machine learning-based model is constructed based on the processed data to forecast the future unaffected values of specific 2 groundwater sensors of 3 types, namely pore pressure, water level in core-drilled hole, and water level in lake.

The raw data includes the measurements from 240 sensors located along the proposed project, which contains not only the three features to be evaluated, but also

the other three supplementary features (temperature, precipitation, air pressure).

Based on the results of data analysis, we decide to use 5 features including the measurement of the sensors to be evaluated, temperature, air pressure, daily precipitation and comprehensive precipitation of one week as inputs for each prediction. Although there are 23 sensors surrounding the project site measuring air pressure, we choose only one specific air pressure sensor as input based on the distances between them and the sensors to be evaluated. Most of the sensors failed to record the continuous measurements of the required time interval, which negatively affects the time analysis, so several interpolation methods are adopted to deal with such discontinuities. A five-column data set with contentious measurements is obtained, which is then turned into a twenty-five-column data set by five-step window rolling. In order to avoid the common phenomenon that the models just learned to simply repeat the previous values, we take the difference between the current measurement and the previous measurement as our predicting target.

To build a long short term memory (LSTM) model, the twenty-five-column data set is further reshaped into a three-dimension tensor (simple sizes*5*5) as input. However, we found that the proposed LSTM model tends to just fit the overall trend. Therefore, another model LightGBM algorithm is compiled to fit the difference between the real measurement and the predicted value of LSTM model. The predicted values of the two models are summed together as the final output.

Specifically, we constructed an univariate model through Prophet algorithm to

further catch the periodic term of ground water level. Then the mean of the predicted values of Prophet and LSTM-LighGBM are taken as the final predicted values of ground water level.