Using ensemble data mining modelling for nonbinary overflow detection in urban flooding

EGU23-9672
EGU General Assembly 2023

Using Ensemble Data Mining Modelling for Nonbinary Overflow Detection in Urban Flooding

Kourosh Behzadian[1], Farzad Piadeh[1], Albert S. Chen[2], Luiza Campos[3], and Zoran Kapelan[4]
[1]University of West London, Engineering, School of Computing and Engineering, London, UK
[2]Centre for Water Systems, Faculty of Environment, Science and Economy, University of Exeter, Exeter, EX4 4QF, UK
[3]Department of Civil, Environmental and Geomatic Engineering, University College London, Gower St, London WC1E 6BT, UK
[4]Department of Water Management, Faculty of Civil Engineering and Geoscience, Delft University of Technology (TU Delft), Delft, Netherlands

Application of data-driven modelling especially using data mining techniques in flood warning systems has received significant attention recently due mainly to its well-explored sustainable solution for alleviating disruptive socio-economic effects of flood occurrence [1]. Various machine learning models with hybrid data mining techniques have been applied for water level prediction or overflow detection. However, the concept of time-series ensemble modelling has yet to be perceived well, particularly application of nonbinary classification for overflow detection and associated flood risk management [2].

This study presents a new real-time nonbinary overflow detection in urban flooding through extraction of rainfall key features by developing weak learner base models and proposing time-series multi-classification ensemble model. This framework is demonstrated by its application on real case study of urban drainage systems (UDS) located in London, UK. Extracted rainfall features which are selected by partial least squares analysis include (1) rainfall duration, (2) rainfall intensity, (3) evidence of previous rainfall occurrence, and (4) rainfall date of the year. These features are then used to develop seven base models including (1) discriminant analysis, (2) decision tree, (3) Gaussian process regression, (4) K-nearest neighbourhood, (5) Naïve bayes, (6) neural network pattern recognition, and (7) support vector machine to detect one of the three condition of (1) overflow, (2) water level rise is expected but drained successfully without any overflow occurrence, (3) no water level rise is expected. A novel ensemble model (ENS) which blends the performance of developed base models into the decision tree structure was then developed for overflow detection of next twelve 15-min timesteps (i.e., 3 hrs). The result performance of this model is compared by two well-practiced models i.e., stacked random forest (ERF), and nagging K-nearest neighbourhood (EKN) [3]. Confusion matrix is selected as a method of performance assessment in which total positive ratio, accuracy, and total negative ratio are picked up as key performance indicators.

Results show two new proposed rainfall features named “evidence of previous rainfall occurrence” and “rainfall date of the year” could significantly enhance the base model’s accuracy. Furthermore, ENS model could reduce overestimation and underestimation miss rates by nearly 10% in total for

3 hrs-ahead overflow detection, whereas these figures are 37% and 39% for total miss rate of ERF and EKN respectively in the same detection duration. Furthermore, the rate of correct high-hazard overflow detections is 88% in comparison to 64% in ERF and 24% in EKN, which highlights superior ability of the proposed model in early warning alarms of high-hazard situations.

References

[1] Rezaie Adaryani, F., Mousavi, S. Jafari, F. (2022). Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN. *Journal of Hydrology*, 614(A), 128463.

[2] Piadeh, F., Behzadian, K., Alani, A. (2022). A critical review of real-time modelling of flood forecasting in urban drainage systems. *Journal of Hydrology*, 607, 127476.

[3] Piadeh, F., Behzadian, K., Alani, A.M. (2022). Multi-Step Flood Forecasting in Urban Drainage Systems Using Time-series Data Mining Techniques. *Water Efficiency Conference*, West Indies, Trinidad and Tobago. repository.uwl.ac.uk/id/eprint/9690 [Accessed 31/12/2022].