A Comparison of Random Forests and Support Vector Machine in River Stage Forecasting

Kun-Hsiang Lin, Hung-Wei Tseng, Chen-Min Kuo, Tao-Chang Yang, Pao-Shan Yu*
Department of Hydraulic and Ocean Engineering, National Cheng Kung University, Taiwan (Republic of China)
No. 1, University Road, Tainan City 70101, Taiwan (Republic of China)
TEL: +886-6-2757575 ext. 63248
E-mail: yups@mail.ncku.edu.tw

This study developed two real-time forecasting models for river stage by using two machine learning algorithms: Random Forests (RF) and Support Vector Machine (SVM). The models use the river stage and rainfall at current and previous time steps as model input to forecast the river stage for the following one hour. For the leading times longer than one hour, the forecasted stage and forecasted rainfall are involved as model input. There are fourteen typhoon events used for model calibration and validation. Nine of them were used for calibration and the rest were used for validation. The model performances were evaluated by six criteria (i.e., root mean squared error, maximum absolute error, correlation coefficient, error of peak stage, error of time to peak, and coefficient of efficiency). The analysis results indicate SVM has a superior performance for one- to two-hour ahead stage forecasting but RF turns out to be better for three-hour ahead. In addition, an apparent time lag is also found in both RF and SVM predictions. The study further examined the models again (i.e., the models with leading time longer than one hour) by using observed data as input instead of forecasted data. The results show an obvious improvement in time lag issue. However, over-shooting estimates are found for both RF and SVM when the river stage is median high. The study suggests including more factors (e.g., wind speed, wind direction, position of typhoon center, etc) for model development may give a better model performance. Moreover, a sensitivity analysis of model parameters and cross-validation process are also suggested for the future work.