H-029

Water Level Prediction for Disaster Management Using Machine Learning Models

Tin Nilar Lin† Hiroshi Watanabe†

***Abstract*— A flood is an overflow of water and becomes the common natural disaster. Prediction of a flood is one of the challenges for disaster management around the world especially in developing countries. Thus, more accurate flood prediction models have been investigated according to the geographical locations. In this paper, we have studied and compared some useful machine learning models such as KNN, SVR and Linear Regression for getting better water level prediction. The proposed approach is applied to Ayeyarwady river in Myanmar. The future water level is predicted based on the time series data of past water levels. By the experiment, KNN (K-Nearest Neighbour) model shows the least mean absolute error and the error rate is just 0.17%. The predicted output of the proposed model agrees in the actual water level. Therefore, KNN model can be the potential solution for successful water level forecasting application in Ayeyarwady river.**

Keywords— water level prediction, time series analysis, KNN, SVR, Linear regression

1. INTRODUCTION

For centuries, human have been challenged by natural disasters and the challenges have been greater today. Among the natural disasters, flood is the common one because of the leading causes of devastation effect on many countries all over the world. Devastation caused by flood is extremely harmful to the social economic life of the people and the country very often especially in developing countries. In most of the developing countries, accurate flood prediction models are not equipped properly. As a result, people from flood prone areas are suffering more from devastating effects of flood than they should. That is why, implementing more accurate and suitable flood prediction model according to geographical locations is desired to resist flood devastations and becomes essential one。

There are several different machine learning models that have been developed in recent years with different objectives and locations for time series weather prediction. Thanh-Tung Nguyen et al. [1] predicted water levels of Mekong River by using three machine learning models LASSO, Random Forest and Support Vector Regression (SVR) and compare the mean absolute error (MAE) and least MAE result for prediction model is 0.486 meter (m). The rainfall prediction model has been developed by using multiple linear regression (MLR) models and computed the Pearson coefficient for five years by M.kannan et al.[2]. Result shows the approximate value of observed and predicted value but does not show the enough accuracy. Kitsuchart Pasupa et al. [3] has applied different machine learning models to forecast the water levels of Chao Phraya River in Thailand and showed the comparison results and the best one. However, they suffer from their own constraints and have uncertainties in prediction depending on regional and local basic. In this work, an effective water level prediction model has been studied and developed with real time data to analyze seasonal water level patterns of Ayeyarwady river.

The paper is organized as follows: Section II describes the study area and data used. Data analysis, comparison and model evaluation are listed in section III. Experimental result is presented in section IV. Finally, section V. concludes the paper.

† — Waseda University, Graduate School of Fundamental Science and Engineering

2. STUDY AREA AND DATA USED

Ayeyarwady river is the largest and vital river of Myanmar and one of the great rivers in Mekong Region. It flows through the country from northern part to the southern part, Adaman sea in the bay of Bengal. It is 2,170 kilo meters long and river's basin is about 413,674 square kilometers covering about 60% of total area of Myanmar and it is the most important commercial waterway.

There are 16 stations for hydrological forecasting along Ayeyarwady river, the main river of Myanmar and often face the local floods especially in rainy seasons and sometimes even in dry seasons due to the influence of climate changes. That is why, we have selected the very first two stations Station-1 (Myintkyienar) and Station-2 (Banmaw) of Ayeyarwady river that is the region of interest. We have collected and analyzed the real-time water level data of these two stations from 2011 to 2016 to train and test the models for predicting five days advance of Station-2 Banmaw station.

3. DATA ANALYSIS, COMPARSION AND MODEL EVALUATION

3.1 Data Analysis

The collected time series data were used for training and validation and testing to develop the proposed model. "Data Normalization" is important in data preprocessing stage. In our study, raw data have been normalized with mean, variance and standard deviation. Experiment results show that normalization with standard deviation is better than mean and variance. Thus, normalization with standard deviation is done for each variable (i.e water level of current day and previous two days of Myintkyienar Station and Banmaw station). The water level is linearly normalized by using the following equation:

$$X_{norm} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[X_{i-}\bar{X}_i]^2} \quad (1)$$

where X_{norm} is the normalized value of the observed variable, N is the number of instances, X_i is the observed values and $\bar{X}_i$ is the mean value of observed variable. After normalization, the error rate of the proposed models can decrease significantly. The Figure

(1) shows the comparison of MAE result before and after normalization of the dataset.

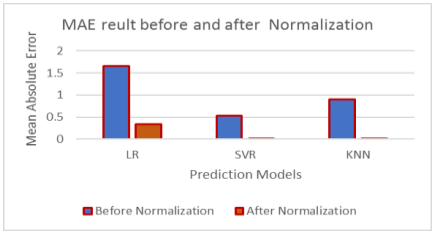

Fig. 1. MAE results before and after normalization

3.2 Model Evaluation

We investigated the dataset by applying different machine learning techniques to evaluate the prediction accuracy in terms of strength and weakness using correlation coefficient, mean absolute error and root mean square error. The acceptable mean absolute error rate (MAE) for flood prediction model of upper Ayeyarwady river is 0.5m ~ 0.7m. In our experimental study, three machine learning models were implemented. These are KNN (K-Nearest Neighbour) [4], Support Vector Regression [1], Linear Regression [2].

In this study, we model water level prediction, so we focus on mean absolute error (MAE). The prediction accuracy can be calculated as follows:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}[|O_i - P_i|] \quad (2)$$

Correlation coefficient (CE) that measures the linear relationship between the actual and predicted values and root mean square error (RMSE) can be used to evaluate the model performance [1].

$$\text{CE} = 1 - \frac{\sum_{i=1}^{N}[O_i - P_i]^2}{\sum_{i=1}^{N}[O_i - \bar{O}_i]^2} \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[O_i - P_i]^2} \quad (4)$$

In these equations, N is the total number of instances, O_i and P_i illustrates observed water levels and predicted water levels and $\bar{O}_i$ is the mean value of observed water levels at i^{th} time respectively.

4. EXPERIMENT RESULTS

In this section, we present the prediction accuracy result using the three machine learning techniques.

As mentioned before, the useful machine learning models including KNN, SVR and Linear regression with 10-fold cross validation have been used to predict the water level of 5 days ahead. Three machine learning models was conducted with 1456 instances and 7 input features.

First, Linear regression predictive model obtained MAE 0.3357 and RMSE 0.4606 whereas CE is 0.7231. Secondly, SVR model showed the better result than Linear Regression and MAE is 0.0048 and RMSE is 0.0063 whereas CE is 1. Finally, the KNN model achieved the best result and MAE is just 0.0017 and RMSE is 0.0097 whereas CE is 0.9999. Table 1 shows the comparison of prediction performance of water level for 5 days ahead. The present KNN model has been worked out for prediction water level of Ayeyarwady River. The actual and predicted water levels of KNN models for 5 days ahead is closely agreement as shown in Figure (2).

Table 1. Water level prediction performance for 5 days ahead in Ayeyarwady river

Model	CE	MAE	RMSE
LR	0.7231	0.3357	0.4606
SVR	**1**	0.0048	**0.0063**
KNN	0.9999	**0.0017**	0.0097

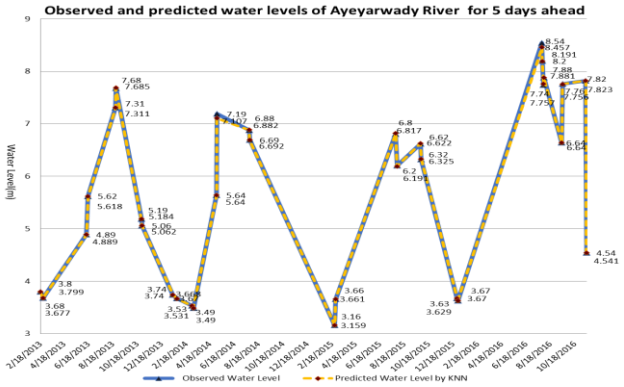

Fig. 2. Observed and Predicted water levels by KNN

5. Conclusion

Water level prediction is the vital factor for flood disaster management. In this paper, three machine learning models were implemented to predict the water level with the best accuracy. KNN model provides the significant accuracy for prediction water level of Ayeyarwady river and the predicted value is also close to actual value. The model also intended to use in the future water level prediction model in Myanmar. Accuracy of our work can be improved by analyzing more time series data and adding useful features to predict the detailed nature of flood. Further, Our model can be extended to predict rainfall, temperature and snow melt etc.

References

[1] Thanh-Tung Nguyen, Quynh Nguyen Huu,Mark Jubjie Li, " Forecasting Time Series Water Levels on Mekong River Using Machine Learning Models," in Seventh International Conference on Knowledge and Systems Engineering, Ho Chi Minh City, Vietnam, 2015, pp.292-297.

[2] M.Kannan, S.Prabhak aran and P.Ramachandran, "Rainfall Forecasting Using Data Mining Technique," International Journal of Engineering and technology, vol. II, no. 6, pp. 397-401, 2010.

[3] Kitsuchart Pasupa, Siripen Jungjareantrat, "Water Levels Forecast In Thailand: A Case Study Of Chao Phraya River," in 14th International Conference on Control, Automation, Robotics & Vision, Phuket, Thailand, 2016, pp.1-6.

[4] Hu L-Y, Huang M-W, Ke S-W, Tsai C-F, “The Distance Function Effect on K-Nearest Neighbor Classification for Medical Datasets.” SpringerPlus 5.1 (2016): 1304. PMC. Web. 16 June 2017.