ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE IN FLOOD FORECASTING: A REVIEW

Azizah Suliman[1], Nursyazana Nazri[1], Marini Othman[1], Marlinda Abdul Malek[1], and Ku Ruhana Ku-Mahamud[2]

[1]Universiti Tenaga Nasional (UNITEN),Malaysia, azizah@uniten.edu.my,nursyazana@uniten.edu.my, marini@uniten.edu.my; marlinda@uniten.edu.my
[2]Universiti Utara Malaysia (UUM), Malaysia, ruhana@uum.edu.my

ABSTRACT. Flood is a natural phenomenon that can cause havocs and deaths. Although flood is sometimes unavoidable, early flood forecasting can be helpful for people to take precaution. In the past decades, researchers have been working on flood forecasting models using artificial intelligence (AI). AI models such as Artificial Neural Network (ANN) and Support Vector Machine (SVM) have been developed and implemented in different locations to help in weather forecasting over the past years. This paper reviews both methods and compares their experimental results.

Keywords: Artificial Neural Network, Support Vector Machine, Flood forecasting, flood

INTRODUCTION

In China, millions of people are affected each year and are forced to evacuate promptly leaving their belongings behind. While in Kuala Lumpur, thousands of people are stranded in the middle of the city center, patiently waiting in their vehicles hoping and praying for it to subside soon. Two different scenarios, but are caused by one same thing - flood. Definition given by Oxford English Dictionary for flood is an overflow of a large amount of water beyond its normal limits. (Abhas et al., 2012) generally characterized flood into fluvial (or river) floods, pluvial (or overland) floods, coastal floods, groundwater floods or the failure of artificial water systems. What causes flood can vary from heavy downpour to sea level rise. It can last for a few hours to days, or even a longer period depending on the cause. The deadliest flood in China that occurred in 1931, also known as 1931 Central China Flood killed 3.7 million people – recorded as the worst case ever. Perhaps, this is the worst natural disaster of 20th century. As time goes by, researchers started to take precaution by developing flood forecasting model in order to give early warning to citizens in order to avoid catastrophe.

FLOOD FORECASTING MODEL

Over the past decades, researchers have shown interest in developing flood forecasting model. Weerts and Beckers (2009) from Netherlands have constructed a framework named Uncertainty Framework for flood and storm surge forecasting. It is built around procedural and operational constraints. The framework is said to help in deciding which method, and in which part of the model chain, it is most suitable to increase the accuracy or quantifying the (predictive) uncertainty of the flood forecast. Figure 1 shows the uncertainty framework that offers a structured approach to reduce the predictive uncertainty.

Uncertainty is divided into three parts in model used for flood forecasting. a) Input boundary conditions for the prediction. b) Initial conditions of the area or model. c) Behavior of the model during the prediction phase.

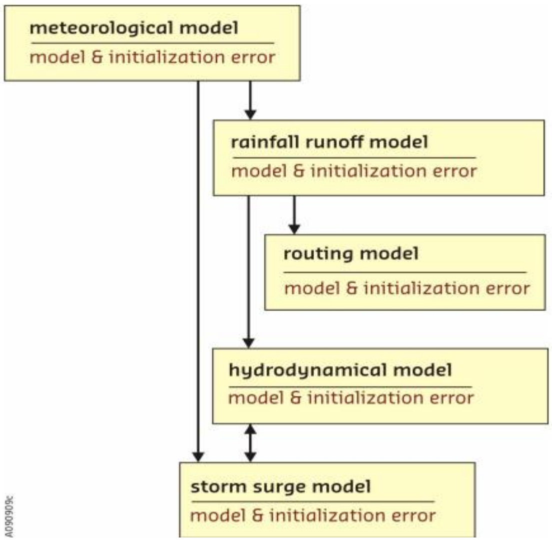

Figure 1. Uncertainty Framework

Although there are other models as stated in the figure that use different algorithms, the application of AI in handling and reasoning under uncertainty has been used widely in diverse areas (Levitt, 1988). The provision of making inferences with uncertainty and the availability of learning mechanism in AI techniques makes it a very useful tool in making prediction and forecasting. Among the common AI method used in flood forecasting are ANN and SVM. This paper focuses solely on models using artificial intelligence and is divided into two parts: (1) Flood forecasting models using Artificial Neural Network; and (2) Flood forecasting models using Support Vector Machine. This paper will further discuss on ANN and SVM in flood forecasting domain in Discussion section.

Artificial Neural Network

ANNs were first introduced to water resources research for their use to predict monthly water consumption and to estimate occurrences of flood. Since then, ANNs have been used for a number of different water resource applications which include time-series prediction for rainfall forecasting, rainfall-runoff processes and river salinity. ANNs have also been used for modeling soil and water table fluctuations, pesticide movement in soils, water table management and water quality management (Parson, 1999).

Models of Artificial Neural Network

Mandal et al. (2005) employed ANN model, namely Multi-layer Perceptron (MLP) using back-propagation network technique and used delta rule for training. Environmental parameters used for this research are temperature, humidity, underground water level, precipitation and wind speed. It is found that underground water level is the most significant parameter for the prediction model. Simulation runs for this model using NeuroSolutions v4.10 has resulted in 97.33% of given overall prediction accuracy.

Ayalew et al. (2007) adopted three-layer back-propagation ANN model for real-time flood forecasting in Omo River, Ethiopia. Floods in Omo river are sudden, non-linear and of short duration. ANN models are best suited for forecasting such types of floods. This research uses sigmoid function which is commonly used for hydrological studies. Two important parameters in this research are magnitude and time-to-peak discharge. Comparisons of observed and forecasted runoff values for training and testing for all models showed little discrepancies.

Tan et al. (2008) combined two models of ANN and SVM in order to come out with a new model called Reward Learning Ensemble (RLEnsemble). One model will learn the problem while the other will learn from the error of its counterpart. SVM is the first model, subsequently followed by ANN using MLP. Error produced by SVM will be the input for MLP. Output produced from MLP will be taken as final prediction. RLEnsemble is the one with highest accuracy in predicting the rainfall pattern in Singapore.

Pang et al. (2011) developed a non-linear perturbation model adopting ANN (NLPM-ANN) and the results are compared to ANN and also linear perturbation model (LPM). In this model, it is recognized that seasonal hydrological behavior, as incorporated in the model is a very important source of information in flood forecasting. It is shown that the NLPM-ANN obtains better simulation results than ANN by 2.7%, while results compared to LPM is higher by 6.32%.

Support Vector Machine

The idea of Support Vector Machine was initially developed in Russia in the 60's by Vapnik and Lerner. Vapnik further developed the field and wrote the definitive book on the subject. A SVM consists of a set of support vectors and a kernel function. The support vectors are a set of vectors from the training data. The support vectors together with the kernel create the function approximation.

Models of Support Vector Machine

Han et al. (2007) is an example that employed Sequential Minimal Optimization algorithm (SMO) for their SVM model. On top of that, they also incorporated an algorithm called SVMLight. The data used for model training is from October 1955 to September 1963, while the testing data is from November 1972 to November 1974. Data are from the catchment in Bird Creek, Oklahoma, USA. Tools used for this research are LIBSVM, coupled with Gunn's Toolbox for data normalization. A comparison with some benchmarking models has been made and it demonstrates that SVM is able to surpass all of them in the test data series, at the expense of a huge amount of time and effort.

Wiriyarattanakul et al. (2008) used fuzzy support vector machine regression (FSVMR) to predict the runoff of Yom River at Sukhotai province, Thailand. They selected runoff data from June until October, between 1995-2000 and 2002-2004. The data are compared using FSVMR and standard SVMR. Average MAE of the best FSVMR model is 3.627 m^3/s and 7.728 m^3/s in the training and testing data set, respectively. While the average MAE of the best SVMR model is 3.954 m^3/s and 8.041 m^3/s in the training and testing data set, respectively. The MAE of the blind test data set from the best FSVMR model and best SVMR model are 7.8588 m^3/s and 9.0895 m^3/s, respectively. This shows that the FSVMR is more effective and efficient in forecasting runoff than the standard SVMR.

Hu et al. (2011) adopted SVM model which provided higher runoff forecast accuracy compared to the forecasts of the ANN model for monthly runoff in the upstream of the Fenhe River. It used a hybrid forecasting technique of support vector regression and its applications for rainfall-runoff forecasting in order to investigate its feasibility in forecasting runoff amounts. Various SVM models were trained to simulate monthly and daily rainfall-runoff relationships and compared with the ANN model. The results show that the SVM model has higher nonlinear mapping capabilities and thus can more easily capture runoff data patterns than can the ANN models.

Bell et al. (2012) adopted SVM for river runoff forecasting, with Smola/Scholkopf's Sequential Minimal Optimization algorithm for training a SVM with a RBF kernel. They used monthly precipitation and snow data gathered from 10 precipitation monitoring stations and 28 snow monitoring stations located in the American River basin. The calculations were made using WEKA v3.6 and the results using SMOreg with a RBF kernel yield a relative absolute error 48.65% versus 63.82% for the human ensemble forecast.

SUMMARY

All the previous works by researchers are summarized as shown in the Table 1 below for easy comparison. Even though a direct comparison might not seem fair as the parameter used differ, it is highly noticeable that SVM does give a better accuracy. (Han et al., 2007, Wiriyarattanakul et al., 2008, Hu et al., 2011, Bell et al., 2012).

Table 1. Summary of ANN and SVM in flood forecasting models

Method		Location	Technique	Parameters	Tools	Outcome
ANN	SVM					
✓		India (Mandal et al., 2005)	Multi-Layer Perceptron (delta rule for training)	i. Temperature ii. Humidity iii. Underground water level iv. Precipitation v. Wind speed	NeuroSolutions v4.10	Water level is the key parameter related to flood. Overall prediction accuracy is 97.33%.
✓		Omo River, Ethiopia (Ayalew et al., 2007)	3 layers back-propagation ANN (sigmoid function)	i. Magnitude ii. Time-to-peak discharge	-	Comparisons of observed and forecasted runoff values for all models showed little discrepancies.
✓	✓	Singapore (Tan et al., 2008)	RLEnsemble (combination of ANN and SVM)	Error produced by SVM will be the input for ANN.	-	RLEnsemble is the one with highest accuracy in predicting rainfall pattern in Singapore.
✓		Lower Yellow River, China (Pang et al., 2011)	Non-linear Perturbation Model adopting ANN	Discharge time series	-	NLPM-ANN obtains better simulation results than the APM and ANN.
	✓	Oklahoma, USA (Han et al., 2007)	SMO, SVMLight	River flow	LIBSVM, Gunn's Toolbox (for data normalization)	It demonstrates that SVM is able to surpass all of the compared models in the test data series.
	✓	Yom River, Thailand	Fuzzy Support Vector Machine	Runoff	-	Average error of FSVMR model is lower than SVMR

		(Wiriyarattanakul et al., 2008)	Regression			models.
	✓	Fenhe River, China (Hu et al., 2011)	Support Vector Regression	Precipitation-Runoff	-	The results show SVM model has higher non-linear mapping capabilities than ANN model.
	✓	American River, California (Bell et al., 2012)	Smola/Scholkopf's Sequential Minimal Optimization	Monthly precipitation and snow data	Machine Learning Tool WEKA	Results using SMOreg yield a relative absolute error of 48.65% versus 63.82% for the human ensemble forecast.

DISCUSSION

G. Zhang et al. (1998) stated reasons why ANN is highly used for forecasting. ANNs are well suited for problems whose solutions require knowledge that is difficult to specify but for which there are enough data or observations. Second, ANNs can generalize. As forecasting is performed via prediction of future behavior from examples of past behavior, it is suitable to be applied in forecasting flood. Records of rainfall in past years can be trained to see the trend and eventually a prediction can be made. Third, ANNs are nonlinear. ANN, which are nonlinear data-driven approaches are capable of performing nonlinear modeling without knowledge about the relationships between input and output variables. Thus, they are a more general and flexible modeling tool for forecasting.

In choosing suitable AI models for forecasting model – not limited to flood, it is always crucial to question ourselves of how well will the model make predictions for events that are not in the training set. As for ANN model, when a little modification is done to be NLPM-ANN model, it becomes a flexible tool for flood forecasting, especially in the area without detailed hydrometer data, a common situation particularly in developing countries. On the other hand, although we can see SVM has been increasingly used in recent hydrological modeling research, it still has its limitations such as poor performance in skewed dataset. Q. Li et al. (2007) stated that SVM is highly dependent on its parameters and the kernel parameters. The inference process of SVM may become time-consuming and computationally expensive due to the large number of support vectors. Looking at the results of previous researches, it is highly recommended to further explore SVM in building flood forecasting model.

CONCLUSION

In reducing side effects of high computation time of SVM, it is also recommended that the use of parallel SVM be investigated. With the availability of GPU and multicore processors on current machines, that would be the best direction to take in the flood forecasting model development. A preliminary work has been done to develop a flood forecasting model for a selected river in Malaysia using parallel SVM on GPU. Further results of this will be discussed in next publication.

ACKNOWLEDGMENT

The authors wish to thank the Ministry of Higher Education Malaysia for funding this study under Long Term Research Grant Scheme (LRGS/b-u/2012/UUM/Teknologi Komunikasi dan Informasi).

REFERENCES

Abhas, K. J., Bloch, R. & Lamond, J. (2012). *Cities and Flooding - A Guide to Integrated Urban Flood Risk Management for the 21st Century*.

Ayalew, L., Möller, D. P. F. & Reik, G. (2007). Using ANN for Real-Time Flood Forecasting. Summer Computer Simulation Conference (SCSC).

Mandal, S., Saha, D., & Banerjee T. (2005). A Neural Network Based Prediction Model for Flood in A Disaster Management System with Sensor Networks. *International Conference on Intelligent Sensing and Information Processing* (ICISIP).

Tan, T. Z., Lee, G. K. K., Liong, S. Y., Lim, T.K., Chu, J. & Hung. T. (2008). Rainfall Intensity Prediction by a Spatial-Temporal Ensemble. *International Symposium on Neural Networks* (ISNN), pp. 1721-1727.

Hu, C., Wu, Z., Wang, J. & Liu L. (2011). Application of the SVM on Precipitation-Runoff Modeling in Fenhe River. *International Symposium on Water Resource and Environmental Protection* (ISWREP).

Bell, B., Wallace, B. & Du. Z. (2012). Forecasting River Runoff through Support Vector Machines. Proc. 11th *IEEE International Conference on Cognitive Informatics & Cognitive Computing* (ICCI*CC'12).

Li, Q., Jiao, L. & Hao Y. (2007). Adaptive Simplification of Solution for Support Vector Machine. *Pattern Recognition*, vol. 40, 972-980.

Wiriyarattanakul, S., Auephanwiriyakul, & Umpon, S. N. T. (2009). Runoff Forecasting Using Fuzzy Support Vector Regression. IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS).

Parson, C. S. (1999). Use of Artificial Neural Networks to Facilitate Hydrological Modeling on the Internet. PhD thesis

Zurada (1992). *Artificial Neural Systems*. West Publishing Company.

Weerts, A. & Beckers, J. (2009), Uncertainty Framework for Operational Flood and Storm Surge Forecasting. *Research for Flood Control 2015 Program*.

Pang, B. & Liang, Y. (2011), Application of an Improved Neural Network to Flood Forecasting of the Lower Yellow River. *4th International Symposium on Computational Intelligence and Design* (ISCID).

Han, D., Chan, & Zhu, L. N. (2007), Flood Forecasting Using Support Vector Machines. *Journal of Hydroinformatics*, IWA Publishing.

Levitt, T. S. (1988), Uncertainty in Artificial Intelligence. *AI Magazine*, vol. 9, Number 4, pp. 77-78.