19th International Computer Science and Engineering Conference (ICSEC)
Chiang Mai, Thailand, 23-26 November, 2015

Using Backpropagation Neural Networks for Flood Forecasting in PhraNakhon Si Ayutthaya, Thailand

Chitsutha Soomlek, Nattawadee Kaewchainam, Thawat Simano, and Chakchai So-In
Department of Computer Science, Faculty of Science, Khon Kaen University
Naimuang, Muang, Khon Kaen, 40002 Thailand
chisutha@kku.ac.th; nattawadee.k@kkumail.com; thawatsimano@gmail.com; chakso@kku.ac.th

***Abstract*—This research created models and an application for predicting the water levels at the gauging stations C.35 located at the Chao Phraya River, PhraNakhon Si Ayutthaya, Thailand by using backpropagation neural networks. The ranges of forecasting are one day, two days, and three days in advance. Rainfall, the water levels, and the water flowing rates in the Chao Phraya River measured from the gauging stations C.2, C.13, C.35, C.36, and C.37 collected in the year 2008-2010 were used for developing the water-level forecasting models. All created models were validated in term of mean square error (MSE), correlation coefficient, and tested in term of model efficiency index and MSE. The best model produced 90.1218% of accuracy and was employed in our flood warning application.**

Keywords—Neural Networks; Prediction; Forecasting; Data Mining; Floods

I. Introduction

For years, Thailand has been suffering from unexpected disastrous floods. In 2011, the worst flooding in the half century of Thailand affected over 13.4 million people, 90 billion square kilometers of land, country's economy, public health, and the estimated damages reached $45 billion [1].

Various approaches had been introduced to relieve and prevent the flood crisis such as flood prevention, flood warning, flood forecasting, water management, etc. [2-7]. Given early warnings may reduce the effect of the natural disaster and could help people prepare for disaster and minimize damages.

The research proposed models and an application for predicting the water levels and, therefore, a possibility of flooding in the central area of Thailand; which contains a number of industrial sectors, agriculture areas, local and international businesses, historical places, tourist attractions, and government services; to relieve the effect of flooding situation in the future. Water-level forecasting models were constructed and evaluated by using backpropagation neural network and various data collected by the regional irrigation office 10 located at Lop Buri province, Thailand [8]. The ranges of forecasting are one day, two days, and three days in advance. The best model was utilized in our flood warning application.

The paper is organized as follows. Section II provides background information and existing work. Section III briefly describes the collected data and the study area. Procedure and methods are explained in section IV. Experimental setup and the results are discussed in section V and VI respectively. The conclusions and future work are also expressed in the last section.

II. Related Work

Neural network (NN) has been applied to various research fields and showed excellent results in water-level forecasting in many regions [5-7]. In addition, there are a number of variations designed to get more accurate results. For example, C. Chen Chang *et al.* developed a Decision Group Back-propagation NN (DBPNN) which was created to improve the traditional BPNN and to solve the problems found in a deterministic model [5]. DBPNN had been combined with various BPNN models to forecast flooding at Wu-Shi watershed in Taiwan [5]. DBPNN could improve reliability and reduced estimation error comparing to the traditional BPNN.

In 2012, a combination of NN and intelligent agents were developed as a flood warning system and designed to analyze the risk of flood caused by rain [6]. In this case, NN was used for forecasting and agents were employed for mass alerts and data collection at certain hydrometeorological stations.

Recently, F. Chang *et al.* employed BPNN, Elman NN, and NARX network for water level forecasting in the area of Taipei City of Taiwan and adopted rainfall collected at six gauging stations and floodwater storage pond water level as inputs of the models [7]. The results indicated that NARX network was better when performing 10-60 minutes ahead forecast. However, one-hour ahead forecast might not give enough time for evacuation and preparation for a disaster.

In this research, BPNN was employed to create models for predicting water-levels in the Chao Phraya River, PhraNakhon Si Ayutthaya, Thailand at one day, two days, and three days in advance. The models were evaluated in term of mean square error (MSE), correlation coefficient, efficiency, and accuracy. The models producing high efficiency and low error were further developed to become a flood warning application.

III. Data and Study Area

Figure 1 shows a map of PhraNakhon Si Ayutthaya province, Thailand and its neighbors. The province is located in the central region of Thailand and is approximately 75 km. from Bangkok [9]. After studying the geographical structure of Thailand, we found that the PhraNakhon Si Ayutthaya province has the highest risk of being flooded, because the area

978-1-4673-7825-3/15/$31.00 ©2015 IEEE

is located in the flat river plain of the Chao Phraya River [9]. According to the map (See Figure1), when the province is getting flooded, there is a high possibility that Bangkok will be flooded as well. Lop Buri and Pa Sak rivers are also met at the PhraNakhon Si Ayutthaya province. When water flows from the northern part of Thailand through the Chao Phraya River is in a large quantity, there is a high possibility that the area will be flooded. Sea water back up from the south and heavy rainfalls are also the causes of flooding in this area.

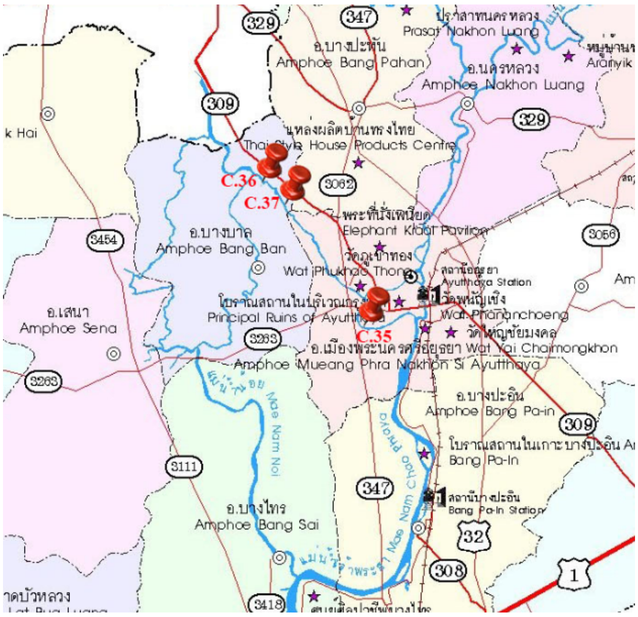

Fig.1. Map of PhraNakhon Si Ayutthaya province, Thailand, figure after [15]

The effects of rainfall, the water levels in the nearby dams, the amount of water releasing from the neighboring areas, and flowing rates were taken into consideration. Therefore, the data collected at the selected gauging stations in the area of study were:

- RC2 – Rainfall at the C.2 water gauging station, in the regional irrigation office, Nakhon Sawan
- RC13 – Rainfall at the C.13 water gauging station, Chao Phraya River (Behind the Chao Phraya Dam at Bang Luang Sub-district, Suppaya District, Chainart)
- RC35 – Rainfall at the C.35 water gauging station, the Chao Phraya River at Ban Pom Sub-district, PhraNakhon Si Ayutthaya District, PhraNakhon Si Ayutthaya
- WLC2 – Water level at the C.2 water gauging station
- WLC13F – Water level in front of the dam at the C.13 water gauging station
- WLC13B – Water level behind the dam at the C.13 water gauging station
- WLC35 – Water level at the C.35 water gauging station
- WLC36 – Water level at the C.36 water gauging station at Bang Ban District, PhraNakhon Si Ayutthaya
- WLC37 – Water level at the C.37 water gauging station at Bang Ban District, PhraNakhon Si Ayutthaya
- QC2 – Flowing rate at the C.2 water gauging station
- QC13 – Flowing rate at the C.13 water gauging station
- QC35 – Flowing rate at the C.35 water gauging station
- QC36 – Flowing rate at the C.36 water gauging station
- QC37 – Flowing rate at the C.37 water gauging station

The data were collected during the year 2008-2010. The C.35 water gauging station is closest to the city center. Normally, it takes approximately six hours for water flowing from the C.36 and C.37 stations to the C.35 station, 24 hours from the C.13 station, and 48 hours from the C.2 station. Please note that August-October is the rainy season in Thailand, therefore, the PhraNakhon Si Ayutthaya province has high risk of being flooded during the season.

IV. Procedure and Methods

A. *Pre-processing*

Since the data was collected by a third-party organization, i.e., the regional irrigation office 10 [8], before the collected data can be used in our experiments, data pre-processing was performed to ensure that the dataset was in their best quality.

At this phase, the data were selected and divided into training dataset and testing dataset. Since the gauging station C.35 is closest to the city center, the water level at the station was considered as the target value. The data collected during August-October 2008-2009 were used as training dataset and the data collected during August-October 2010 were used as testing dataset. There were 278 records of data all together.

After that, the correlations among the rainfall, water levels, and flowing rates at the gauging station C.2, C.13, C.36, and C.37 and the water level at one day, two days, and three days before the chosen date at C.35 station were evaluated using Pearson product-moment correlation coefficient (r). The results are presented in Table I.

TABLE I. CORRELATION COEFFICIENTS

Correlation with the water level at C.35 Station: One day ahead					
Station, Chosen date (t)	**r**	**Station, One day ahead (t-1)**	**r**	**Station, Two days ahead (t-2)**	**r**
RC35	-0.001	RC13	0.100	RC2	0.106
WLC35	0.994	WLC13F	0.277	WLC2	0.956
WLC36	0.987	WLC13B	0.060	QC2	0.955
WLC37	0.987	QC13	0.976		
QC35	0.980				
QC36	0.977				
QC37	0.923				
Correlation with the water level at C.35 Station: Two days ahead					
Station, Chosen date (t)	**r**	**Station, One day ahead (t-1)**	**r**	**Station, Two days ahead (t-2)**	**r**
RC35	0.004	RC13	0.131	RC2	0.142
WLC35	0.946	WLC13F	0.256	WLC2	0.898
WLC36	0.933	WLC13B	0.056	QC2	0.898
WLC37	0.938	QC13	0.928		
QC35	0.924				
QC36	0.919				
QC37	0.886				
Correlation with the water level at C.35 Station: Three days ahead					
Station, Chosen date (t)	**r**	**Station, One day ahead (t-1)**	**r**	**Station, Two days ahead (t-2)**	**r**
RC35	0.004	RC13	0.170	RC2	0.164
WLC35	0.897	WLC13F	0.193	WLC2	0.846
WLC36	0.887	WLC13B	0.051	QC2	0.839
WLC37	0.892	QC13	0.865		
QC35	0.875				
QC36	0.856				
QC37	0.814				

B. Model formulation

According to the correlation coefficients shown in Table I, the prediction functions of the water levels at the C.35 station can be formulates as follows:

WLC35(t+1) = f{WLC35(t), WLC36(t), WLC37(t), QC35(t), QC36(t), QC37(t), WLC2(t-2), QC2(t-2), QC13(t-1)} (1)

WLC35(t+2) = f{WLC35(t), WLC36(t), WLC37(t), QC35(t), QC36(t), QC37(t), WLC2(t-2), QC2(t-2), QC13(t-1)} (2)

WLC35(t+3) = f{WLC35(t), WLC36(t), WLC37(t), QC35(t), QC36(t), QC37(t), WLC2(t-2), QC2(t-2), QC13(t-1)} (3)

Where, t is the chosen date, WLC35(t+1) predicts the water level at C.35 gauging station one day ahead, WLC35(t+2) predicts the water level at C.35 gauging station two days ahead, and WLC35(t+3) predicts the water level at C.35 gauging station three day ahead.

The results presented in Table I indicated that the one-day ahead water level at the C.35 gauging station is highly correlated to WLC35(t), WLC36(t), WLC37(t), QC35(t), QC36(t), QC37(t), WLC2(t-2), QC2(t-2), and QC13(t-1). In case of two-day and three-day ahead water levels, the water levels at the gauging station have high correlation to the same factors but lower than the one-day ahead water level. Therefore, there is a high possibility that WLC35(t+1) will produce the highest accuracy when using the same input data as WLC35(t+2) and WLC35(t+3).

In this research, the criteria proposed in [10] were employed to calculate the number of nodes in the hidden layer of a neural network. The structures of the candidate neural networks are as follows:

	Input nodes-Hidden nodes-Output node
WLC35(t+1)	9-5-1, 9-7-1, 9-17-1, 14-8-1, 14-10-1, 14-27-1, 9-5-5-1, 9-5-6-1, 9-5-7-1, 9-5-8-1, 9-5-9-1, 9-5-10-1, 9-5-11-1, 9-5-12-1, 9-5-13-1, 9-5-14-1, 9-5-15-1, 9-5-16-1, 9-5-17-1, 14-8-8-1, 14-8-9-1, 14-8-10-1, 14-8-11-1, 14-8-12-1, 14-8-13-1, 14-8-14-1, 14-8-15-1, 14-8-16-1, 14-8-17-1, 14-8-18-1, 14-8-19-1, 14-8-20-1, 14-8-21-1, 14-8-22-1, 14-8-23-1, 14-8-24-1, 14-8-25-1, 14-8-26-1, 14-8-27-1
WLC35(t+2)	9-5-1, 9-7-1, 9-17-1, 14-8-1, 14-10-1, 14-27-1, 9-5-5-1, 9-5-6-1, 9-5-7-1, 9-5-8-1, 9-5-9-1, 9-5-10-1, 9-5-11-1, 9-5-12-1, 9-5-13-1, 9-5-14-1, 9-5-15-1, 9-5-16-1, 9-5-17-1, 14-8-8-1, 14-8-9-1, 14-8-10-1, 14-8-11-1, 14-8-12-1, 14-8-13-1, 14-8-14-1, 14-8-15-1, 14-8-16-1, 14-8-17-1, 14-8-18-1, 14-8-19-1, 14-8-20-1, 14-8-21-1, 14-8-22-1, 14-8-23-1, 14-8-24-1, 14-8-25-1, 14-8-26-1, 14-8-27-1
WLC35(t+3)	9-5-1, 9-7-1, 9-17-1, 14-8-1, 14-10-1, 14-27-1, 9-5-5-1, 9-5-6-1, 9-5-7-1, 9-5-8-1, 9-5-9-1, 9-5-10-1, 9-5-11-1, 9-5-12-1, 9-5-13-1, 9-5-14-1, 9-5-15-1, 9-5-16-1, 9-5-17-1, 14-8-8-1, 14-8-9-1, 14-8-10-1, 14-8-11-1, 14-8-12-1, 14-8-13-1, 14-8-14-1, 14-8-15-1, 14-8-16-1, 14-8-17-1, 14-8-18-1, 14-8-19-1, 14-8-20-1, 14-8-21-1, 14-8-22-1, 14-8-23-1, 14-8-24-1, 14-8-25-1, 14-8-26-1, 14-8-27-1

Please note that 9-input BPNNs used WLC35(t), WLC36(t), WLC37(t), QC35(t), QC36(t), QC37(t), WLC2(t-2), QC2(t-2), and QC13(t-1) as their inputs. In case of 14-input BPNNs, RC35(t), WLC35(t), WLC36(t), WLC37(t), QC35(t), QC36(t), QC37(t), RC13(t-1), WLC13F(t-1), WLC13B(t-1), QC13(t-1), RC2(t-2), WLC2(t-2), and QC2(t-2) were the inputs of the neural networks.

C. Validation and Testing

The models were validated in term of MSE, r, and tested in term of model efficiency index (EI) and MSE. The selected data were normalized before being used in the experiments. Two approaches were conducted in the validation process:

- Case1: 70% of the data collected during August-October 2008-2009 were used as training dataset. 15% were used for validation and another 15% were used for testing purposes.
- Case2: *k*-fold cross-validation technique [11] was employed to validate the models. In this case, the data collected during August-October 2008-2009 were used.

After the models were validated, BPNNs having high correlation (r>0.6) and MSE closed to 0 were tested by comparing their prediction results with the actual water levels. The data collected during August-October 2010 were used to measure the efficiency and the accuracy of the models, i.e., testing dataset. BPNNs producing best results will be employed in our flood warning application.

V. EXPERIMENTAL SETUP

For Case1 in the model validation process (explained in the previous section), the models were developed in MATLAB by employing nnstart toolbox, i.e., a neural network toolbox [12], and trainlm, i.e., a Levenberg-Marquardt neural network training function [13]. The maximum number of training rounds was 1000 epochs. As for Case2, Weka 3 [14] was utilized.

VI. EXPERIMENTAL RESULTS

This section presents the selected results obtained from the training phase, model validation, and forecast performance of the BPNNs. The results and discussions are explained in details as follows.

Table II presents the selected results obtained from the one-hidden-layer BPNNs in the training phase by utilizing the data prepared in Case1. In case of the two-hidden-layer BPNNs, the MSE produced by the networks were relatively high.

TABLE II. RESULTS OBTAINED FROM THE TRAINING PHASE

Model	Structure of the BPNN	Training		Validation		Testing	
		MSE	r	MSE	r	MSE	r
WLC35 (t+1)	9-5-1	0.0191	0.9944	0.0151	0.9958	0.0052	0.9964
	9-7-1	0.0101	0.9958	0.0011	0.9960	0.0014	0.9962
	14-27-1	0.0169	0.9964	0.0604	0.9870	0.0186	0.9958
WLC35 (t+2)	9-7-1	0.1554	0.9536	0.1085	0.9772	0.1276	0.9697
	14-8-1	0.1317	0.9619	0.1036	0.9801	0.1024	0.9711
	14-10-1	0.1397	0.9673	0.1031	0.9725	0.1450	0.9735
WLC35 (t+3)	9-5-1	0.2139	0.9362	0.2823	0.9246	0.1365	0.9561
	14-10-1	0.1221	0.9691	0.1389	0.9568	0.1515	0.9562
	14-27-1	0.1381	0.9679	0.2324	0.9432	0.1830	0.9562

There was a possibility that dividing the training data and testing data by percentage affected the quality of the model. Therefore, *k*-fold cross-validation technique [11] was employed in our model validation process. Table III presents the validation results.

TABLE III. RESULTS OBTAINED FROM K-FOLD CROSS-VALIDATION

Model	Structure of BPNNs	Fold	RMSE	r
WLC35(t+1)	9-5-1	7	0.1441	0.9961
	9-7-1	7	0.1442	0.9961
	14-27-1	10	0.4968	0.9547
WLC35(t+2)	9-7-1	4	0.3993	0.9727
	14-8-1	7	0.4215	0.9675
	14-10-1	6	0.4236	0.9670
WLC35(t+3)	9-5-1	10	0.4983	0.9544
	14-10-1	6	0.5012	0.9534
	14-27-1	9	0.4869	0.9556

The results from Table II and III confirmed that the 9-5-1 and 9-7-1 BPNNs should be employed for predicting water level one day in advance since both netwoks produced high correlation and low MSE in the validation process. In addition, there was no significant different in the correlation and MSE values produced by both BPNNs. In case of the two-day ahead prediction, 14-10-1 BPNN had the best results when using the data prepared for validation in Case1; however, 9-7-1 BPNN provided the best results when using *k*-fold cross-validation. For the three-day ahead water-level prediction, 14-10-1 and 14-27-1 BPNNs produced the best results when using data prepared in Case1 and Case2 respectively.

TABLE IV. EFFICIENCY INDEX AND MEAN SQUARE ERROR OBTAINED FROM THE TESTING PHASE

Model	Structure of the BPNN	EI	MSE
WLC35(t+1)	9-5-1	0.0185	1.1325
	9-7-1	0.9012	0.1139
	14-27-1	-0.6254	1.8755
WLC35(t+2)	9-7-1	-4.2021	5.8942
	14-8-1	0.2646	0.8331
	14-10-1	0.5595	0.4990
WLC35(t+3)	9-5-1	-6.8828	8.8036
	14-10-1	0.3833	0.6887
	14-27-1	-3.5051	5.0315

At the testing phase, in order to find the efficiency and accuracy of the water-level forecasting models, the data collected during August-October 2010 were employed as input data. Table IV shows the results obtained from the testing phase. The results indicated that the best model for forecasting the water level of the Chao Phraya River at the C.35 gauging station in PhraNakhon Si Ayutthaya province was the one-day ahead forecasting model when predicting with the 9-7-1 BPNN (EI = 0.9012, MSE= 0.1139). The prediction results of the model are illustrated in Figure2. Red dash line represents the actual water level at the C.35 gauging station and the blue solid line presents the predicted water level. The accuracy of the model was 90.1218%.

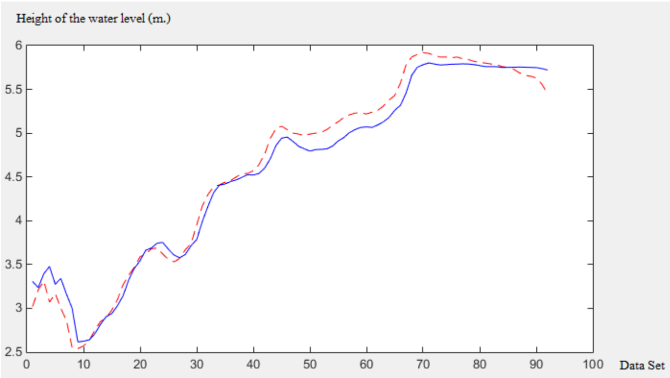

Fig.2. Results produced by the one-day ahead water level forecasting model (BPNN = 9-7-1)

Fig.3. The prototype of the flood warning application

Figure 3 presents the prototype of the flood warning application. Since the 9-7-1 BPNN produced the most reliable results, it was employed in the flood warning application. When the predicted water level is higher than the height of river bank, the application will give a warning message and approximate elevated water level, i.e., predicted water level minus the height of river bank, to a user. Currently, the application can predict the water level at one day in advance only since the other water-level forecasting models produced poor accuracy; therefore, they were not employed in the application.

VII. Conclusions and Future Work

Flooding is a natural disaster that leaves severe impairments to people, economy, and society in various parts of the world, including Thailand. The causes of flood crisis range from natural phenomena to manmade. Thus, this research introduced water-level forecasting models that can predict the water-levels at the C.35 gauging station in the Chao Phraya River, PhraNakhon Si Ayutthaya, Thailand. The models forecast the water levels one day, two days, and three days in advance. The model producing the highest accuracy was further developed to become a flood warning application to relieve the severity of damages caused by the natural disaster.

The data collected at five gauging stations located in and nearby PhraNakhon Si Ayutthaya, Thailand were employed to train and test the water-level forecasting models. The accuracy of the model designed for predicting the water level at the C.35 gauging station one day in advance was 90.1218%.

In order to improve the accuracy of the water-level forecasting model, further investigation is required. Variations of ANN could be employed to find a better model. More training dataset and testing dataset may give us more effective results. These are left for future work.

Acknowledgment

This research was supported by Khon Kaen University, Thailand. We greatly appreciate the research funding from the Khon Kaen University. We would also like to show our gratitude to the regional irrigation office 10 located at Lop Buri province, Thailand for collecting and providing data to this research.

References

[1] "2011 Thailand Flood Executive Summary," Thai Integrated Water Resource Management, 2012. [Online]. Available: http://www.thaiwater.net/web/index.php/ourworks2554/379-2011flood-summary.html. [Accessed: 14 May 2015]

[2] B. Srikudkao, T. Khundate, C. So-In, P. Horkaew, C. Phaudphut, and K. Rujirakul, "Flood Warning and Management Schemes with Drone Emulator Using Ultrasonic and Image Processing," *Adv. in Intell. Syst. and Comput.*, vol. 361, pp. 107-116, 2015.

[3] C. H.R. Lima, U. Lall, T. J. Troy, and N. Devineni, "A Climate Informed Model for Nonstationary Flood Risk Prediction: Application to Negro River at Manaus, Amazonia," *J. of Hydrology*, vol. 522, pp. 594-602, 2015.

[4] C. Damle and A. Yalcin, "Flood prediction using Time Series Data Mining," *J. of Hydrology*, vol. 333, pp. 305-316, 2007.

[5] C. Chen, B. P. Chen, F. N. Chou, and C. Yang, "Development and application of a decision group Back-Propagation Neural Network for flood forecasting," *J. of Hydrology*, vol. 385, pp. 173-182, 2010.

[6] V. Lopez, S. Medina, and J. F. de Paz, "Taranis: Neural networks and intelligent agents in the early warning against floods," *J. of Expert Syst. With Applicat.*, vol. 39, pp. 10031-10037, 2012.

[7] F. Chang, P. Chen, Y. Lu, E. Huang, and K. Chang, "*Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control*," *J. of Hydrology*, vol. 517, pp. 836-846, 2014.

[8] "Regional Irrigation Office 10, Royal Irrigation Department, Lop Buri" [Online]. Available: http://irrigation.rid.go.th/rid10/web%20Eng%2010/index_2.htm. [Accessed: 14 May 2015]

[9] S. Charungthanakij, "An integrated gis-based modeling for sustainable industrial-agricultural land-use planning: case study of Phra Nakhon Si Ayutthaya province," Ph.D. dissertation, Geoinformatics, Suranaree Univ. of Tech., Nakornratchasima, Thailand, 2010.

[10] P. Anusornteerakul, "Thailand's gold price forecast using neural network," Master thesis, Dept. of Comp. Sci., Khon Kaen Univ., Khon Kaen, Thailand, 2009.

[11] D. Larose and C. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, 2nd ed. Hoboken: John Wiley & Sons, Inc., 2014, pp. 139-141.

[12] "nnstart: Neural network getting started GUI" [Online]. Available: http://www.mathworks.com/help/nnet/ref/nnstart.html. [Accessed: 11 March 2015]

[13] "trainlm: Levenberg-Marquardt backpropagation" [Online]. Available: http://www.mathworks.com/help/nnet/ref/trainlm.html. [Accessed: 11 March 2015]

[14] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, "The WEKA Data Mining Software: An Update," SIGKDD Explorations, vol. 11, no. 1, 2009.

[15] "Phra Nakhon Si Ayutthaya Tourist Map" [Online]. Available: http://www.oceansmile.com/map/k/3149.gif [Accessed: 23 December 2014]