Predictive Mining of Rainfall Predictions Using Artificial Neural Networks for Chao Phraya River

Norraseth Chantasut, Charoen Charoenjit, and Chularat Tanprasert

Data Warehouse Technology Section,
National Electronics and Computer Technology Center, Thailand
{norraseth.chantasut, charoen.charoenjit, chularat.tanprasert}@nectec.or.th

Abstract - The rainfall is one of the significant data set of water resource management. With the monthly historical rainfall data in the period of 1941-1999 form 245 rainfall monitor stations in Thailand around Chao Phraya River, the rainfall prediction with an artificial intelligent technique is possible. Artificial neural networks is one the most widely supervised techniques of data mining. It can be applied on predictive mining tasks to make a prediction. The main contribution of this paper is to utilize a neural network model for monthly rainfall prediction. The training and testing patterns are prepared as a time-series data of the past ten months. The numbers of training and testing patterns are 372 and 96, respectively. In the training step, the neural network gives 99.6 % of accuracy and 96.9 % of accuracy in the testing step. The results show that it is possible to predict annual rainfall one year ahead with acceptably accuracy.

Keywords - Neural networks, Rainfall prediction, Data mining

1. Introduction

In Asia region, Thailand has many water-related problems such as water flood and drought. The necessary of water resources management in Thailand is how to acquire useful information for decision making and planning. Actually, the historical rainfall data in the period of 1941-1999 from 245 rainfall monitor stations in Thailand is available for data mining. Thus, the data mining techniques for water resources management is helpful to predict rainfall quantitatively which help in crop planting decisions and reservoir water resource allocation in Thailand.

Artificial Neural Networks (ANNs) has been increasingly applied in various aspects of science and engineering because of its ability to model both linear and non-linear systems without the need to make assumption as are implicit in most traditional statistical approaches. For hydrological modeling problems, ANNs have been used in forecasting model rainfall prediction. (S. Lee, S. Cho and P.M. Wong, 1998) had proposed a divide-conquer approach to divide the region into four sub-area and each is modeled with a different method. Predictions in two larger areas were made by neural networks and predictions in two smaller areas were made by a simple linear regression model. Comparison with the observed data revealed that the artificial neural networks produced good predictions while the linear models produced poor predictions. (M. Chayanis, Oki Taikan and Kanae Shinjiro, 2003) worked with the prediction of monthly rainfall at Chiangmai station in Chao Phraya river basin using neural networks. A sixteen rainfall monitoring stations in Chao Paraya river basin, the Sea Surface Temperature (SST) areas around Thailand and the Southern Oscillation Index (SOI) were employed as the predictors. In an additional study, ANNs have been used in forecasting model of Chao Phraya river flood levels in Bangkok (Tawatchai Tingsanchali, 2000), neural network models for river flow forecasting (Nguyen T. Danh, Huynh N. Phien and Ashim D. Gupta,1999). However, only few applications on rainfall prediction have been reported.

In this paper, the monthly rainfall data in the period of 1941-1999 from 245 rainfall monitor stations in Thailand are archived quantitatively. The quantitative prediction of monthly rainfall in Thailand by backpropagation neural network is examined.

This paper is reprinted from a paper published in Proceedings of Joint Conference The 4th International Conference of The Asian Federation of Information Technology in Agriculture and The 2nd World Congress on Computers in Agriculture and Natural Resources, August 9-12, 2004, Bangkok, Thailand, pp. 117-122.

2. Methodology

Predictive mining is a task that it performs inference on the current data in order to make a prediction. A monthly rainfall data can be grouped as a time-series set because it consists of sequences of values in time. A time-series data can be simplified notation as

$$y = f(t) \tag{1}$$

Where *y* can be any single valued variable which develops in time *t*, in this work, *y* is a monthly rainfall values. To forecast time-series data, it involves knowing the past history of *f* and extrapolating it to the future. The characteristic of the forecasting model is non-linear system, so that the backpropagation neural network can be applied in time-series prediction areas. The Stuttgart Neural Network Simulator (SNNS, http://www-ra.informatik.uni-tuebingen.de/SNNS/) was used to perform the neural network modeling operations to construct a network for monthly rainfall prediction.

3. Data preprocessing

A monthly rainfall data in the period of 1941-1999 from 245 rainfall monitor stations in Thailand were collected from Thailand Integrated Water Resource Management System via the http://www.thaiwater.net/, we have performed some data preprocessing steps on raw set of monthly rainfall data as shown below:

1) Firstly, a monthly rainfall data were cleaned by filling in missing values with mean values.

2) Secondly, a monthly rainfall data were normalized by a min-max normalization into a specified range 0.0 to 1.0

$$NORM_V = \frac{v - \min_A}{\max_A - \min_A} * (new\_\max_A - new\_\min_A) + new\_\min_A \tag{2}$$

Where NORMV is the normalized data, v is the original data of attribute A, $\max_A$ is the maximum values of an attribute A, $\min_A$ is the minimum values of an attribute A. A min-max normalization maps a value v of attribute A to $NORM_V$ in the range [new_min_A, new_max_A]. In this work, we set the new_min_A to 0.0 and new_max_A to 1.0.

4. Neural network architecture

The neural network's weight is initialized by random values between -1.0 to 1.0 and a three layer feed-forward neural network architecture was created (Fig.1, 2). The data were processed into 11 variables: SUM(RAIN) [t], SUM(RAIN) [t-1 to t-10].

The input nodes correspond to summary of monthly rainfall over 245 stations for the past ten months, SUM(RAIN) [t-1 to t-10]. The output node was for summary of current monthly rainfall, SUM(RAIN) [t]. So, the number of input nodes is equal to the number of input features of training examples. In this work, the architecture of neural network in this research is 10:5:1 (input node: hidden node: output node).

Some of training examples and testing examples are shown in Table 1 and Table 2, respectively. The format of input pattern for SNNS consists of input features and output features. For example, the input feature of the first input pattern consists of the summary of rainfall values in January from year 1941 to year 1950 and the output feature of the first input pattern consists of the summary of rainfall values in January in year 1951. The input feature of the second input pattern consists of summary of rainfall values in February from year 1941 to year 1950 and the output feature of the second input pattern consists of the summary of rainfall values in February in year 1951. This process was continued until we have obtained all 372 training patterns. The same process is also applied to the testing patterns set.

Table 1: Example of the training data from 1941 to 1981

No	Month/Year	Sum(Rainfall)	Normalized
1	1/1941	622.3	0.01100
2	2/1941	585.5	0.01035
...	...	...	...
372		1083.5	0.01916

Table 2: Example of the testing data from 1992 to 1999

No	Month/Year	Sum(Rainfall)	Normalized
1	1/1982	251.7	0.00445
2	2/1982	229.1	0.00405
...	...	...	...
96		0.0	0.00000

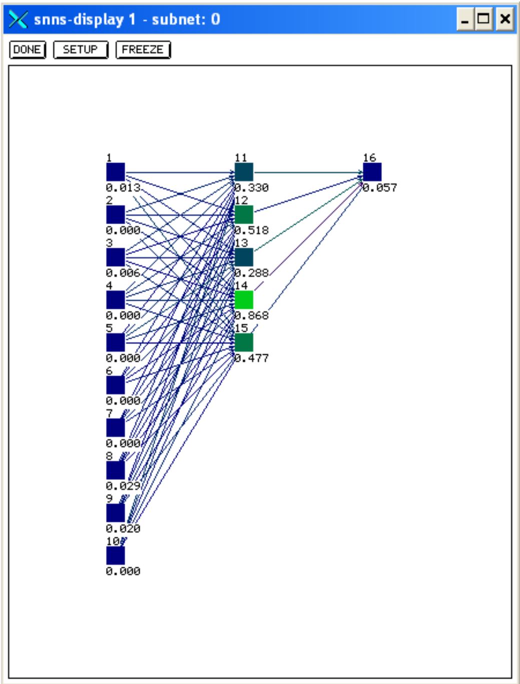

Figure 1. 2D Neural Network Architecture

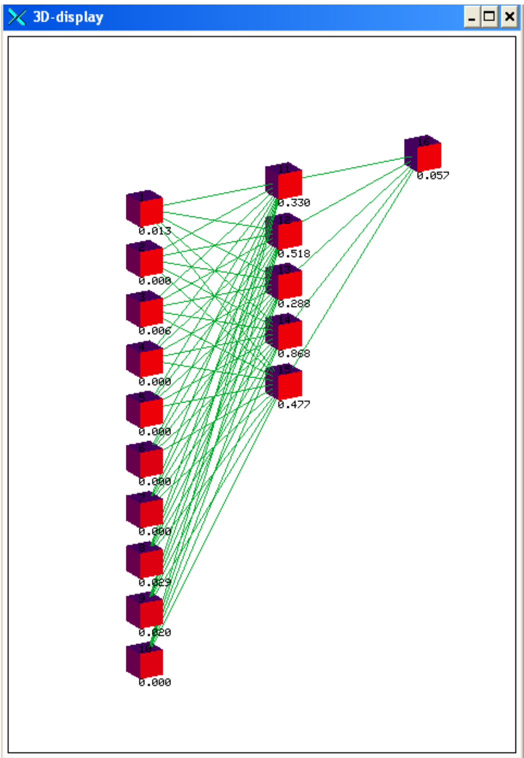

Figure 2. 3D Neural Network Architecture

5. Estimation of Accuracy

To estimate accuracy of a prediction (Tom M. Mitchell 1997), the accuracy is defined as

$$accuracy = \left(1 - \left[\frac{\frac{1}{2}\sum_{i=1}^{N}(t_i - o_i)^2}{N}\right]\right) * 100 \qquad (3)$$

Where t_i is the target output for training example i, o_i is the output of the considered unit for training example i, and N is the number of all training examples. The accuracy values are in the interval [0 %-100 %] and larger accuracy values indicate higher accuracy quality.

6. Experimental Results

The Backpropagation neural network provided by Stuttgart Neural Network Simulator (SNNS) was used in this work. The established ANN model has the accuracy of 99.6 % in the training step. In the testing step, all set of parameters obtained from training step were applied directly, consequently, less accuracy was obtained. However, its accuracy is still tolerable with 96.9 % of efficiency (Table 3). Moreover, during the period in 1992-1994, the given value of Mean Square Error is 0.00486 and the given value of accuracy measure is 99.51 %. It was found that the predicted results are good compared with the observations (Fig. 3).

7. Conclusion and Future Works

For rainfall prediction, artificial neural network was applied to predict the summary rainfall data in Thailand. According to the experiments, predictions of the summary rainfall data using backpropagation neural network were acceptably accuracy. In the future works, some additional inputs were employed for rainfall prediction such as Sea Surface Temperature (SST) areas around Thailand and Southern Oscillation Index (SOI).

Table 3: Accuracy of summary rainfall prediction (network architecture 10:5:1)

Period	MSE	Accuracy
Training (1941-1981)	0.004000	99.6 %
Testing (1982-1999)	0.030799	96.9 %

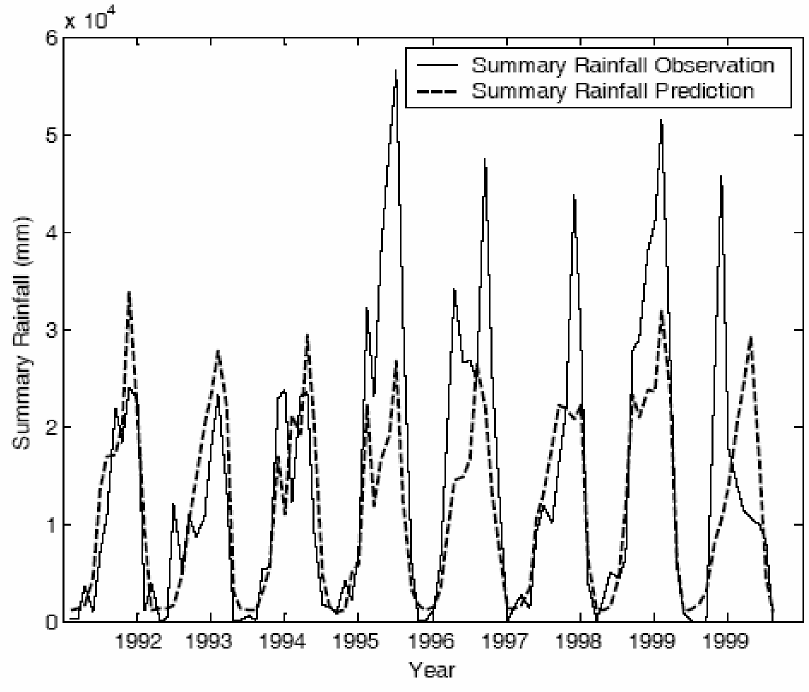

Figure 3. Summary rainfall prediction on testing patterns (1982-1999)

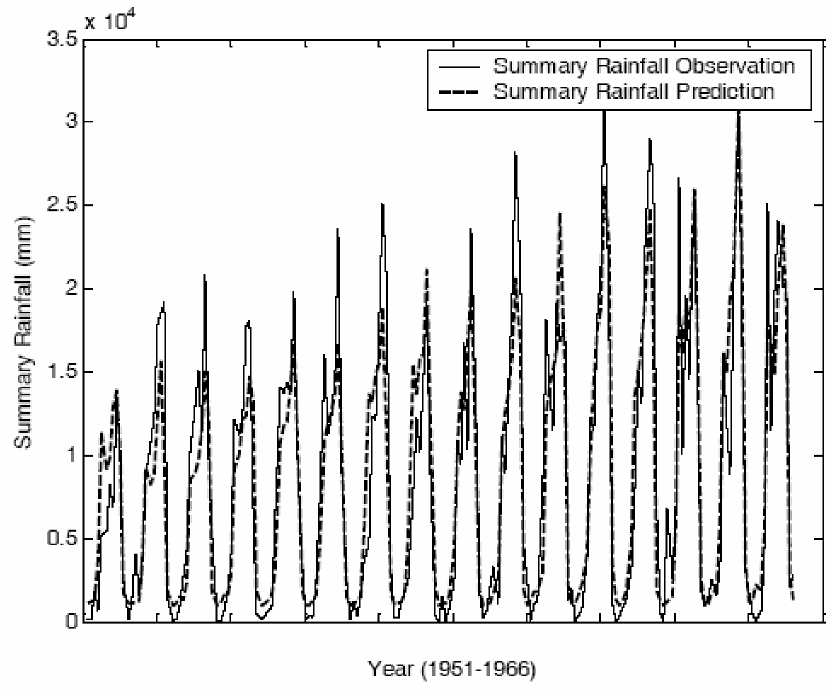

Figure 4. Summary rainfall prediction on training patterns (1941-1966)

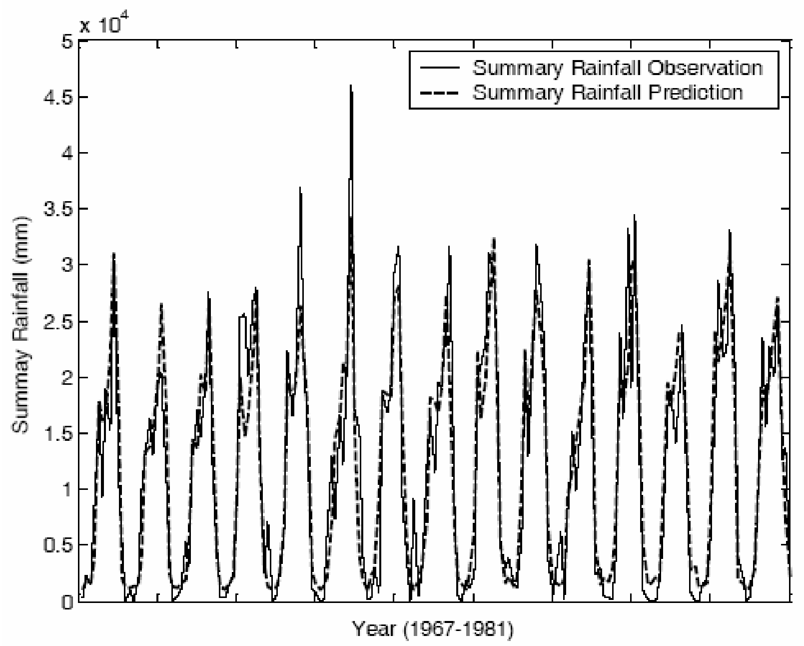

Figure 5. Summary rainfall prediction on training patterns (1967-1981)

Acknowledgement

We extend thanks to Thailand Integrated Water Resource Management System (TIWRM), http://tiwrm.hpcc.nectec.or.th or http://www.thaiwater.net/, for providing related data. This research was performed at the National Electronics and Computer Technology Center, THAILAND.

References

[1] Manusthiparom Chayanis, Oki Taikan, Kanae Shinjiro 2003, "Quantitative Rainfall Prediction in Thailand," International Conference on Hydrology and Water Resources in Asia Pacific. Kyoto. Japan.

[2] Tawatchai Tingsanchali 2000, "Forecasting model of Chao Phraya river flood levels at Bangkok," International Conference on Chao Phraya Delta. Bangkok. Thailand.

[3] Nguyen T. Danh, Huynh N. Phien and Ashim D. Gupta 1999, "Neural network models for river flow forecasting," Journal of Water SA, Vol. 25.

[4] S.Lee, S. Cho and P.M. Wong 1998, "Rainfall Prediction using Artificial Neural Networks," Journal of Geographic Information and Decision Analysis, Vol. 2, No.2, pp. 233-242

[5] SNNS, http://www-ra.informatik.uni-tuebingen.de/SNNS/

[6] Tom M. Mitchell 1997, "Machine Learning," Mcgraw-Hill Press International Editions, pp. 97

Norraseth Chantasut received his M.Sc. degree in Information Technology (Information Science) from King Mongkut's Institute of Technology Ladkrabang, Thailand. in 2004. He received his bachelor degree in Computer Science from Rajabhat Institute Mahasarakham, Thailand in 1999. Currently, Norraseth Chantasut is a research assistant in Data Warehouse Technology Section, Computing Research and Development Division, National Electronics and Computer Technology Center, Thailand. His research interests include document clustering, neural networks and data mining.

Charoen Charoenjit received his B.S. in Tech. Ed. (Computer Technology) in 2000 from the King Mongkut Institute of Technology (North Bangkok) Thailand. During 1999 - 2000 he was a member of the total solution team solving the y2k problem at R&D Computer System Co.,Ltd. In 2001 Charoen Charoenjit has joined the National Electronics and Computer Technology Center (NECTEC) as research assistant. His research interests in data mining.

Chularat Tanprasert received her BS (Mathematics 1st honor, 1989) from Chulalongkorn University, and her MS (1991) and Ph.D. (1994) degrees in Computer Science from University of Louisiana at Lafayette, USA. After her graduation, she has joined the Software and Language Engineering Laboratory (SLL) of the National Electronics and Computer Technology Center (NECTEC) and worked in Thai OCR project until year 2000. Now she is working in the Computing Research and Development Division (RDC) of NECTEC. Her main responsibility is to do research on data warehouses and data mining with the emphasis on knowledge management and bioinformatics. She has also been involved in a number of practical projects such as the development of Thai word processor, web 13, and the Thai speaker identification. She is currently an editor-in-chief of NECTEC Technical Journal since 1999. Her mission as a computer scientist is to advance the theory and application of information processing and computer technologies. Her areas of interest include neural networks, pattern recognition, database systems, artificial intelligence, data mining, data warehouses, bioinformatics, and knowledge management.