Proceedings of the International MultiConference of Engineers and Computer Scientists 2010 Vol I,
IMECS 2010, March 17 - 19, 2010, Hong Kong

A Neuro-Fuzzy Approach for Daily Rainfall Prediction over the Central Region of Thailand

Pramote Luenam, Supawadee Ingsriswang, Lily Ingsrisawang,
Prasert Aungsuratana, and Warawut Khantiyanan

***Abstract*— The methodology of neuro-fuzzy will be presented for the rain forecasting system over the central region of Thailand. The neuro-fuzzy approach was applied to create a classifier for rain prediction. The objective of this work is to demonstrate what relationship models between rain occurrence and other weather features can be developed for predicting accurate rainfall estimates to support the decisions to launch cloud seeding operations. Data sets were collected during 2004 to 2008 from the Chalermprakiat Royal Rain Making Research Center at Hua Hin, Prachuap Khiri khan, and Thai Meteorological Department. A total of 179 records with 57 features were merged and matched by unique date. There are three main parts in this work. Firstly, a correlated-based feature selection (CFS) was used to evaluate the most important features for rain prediction and rainfall level classification. Secondly, a neuro-fuzzy algorithm, NEFCLASS, was used for prediction of rain or no-rain events. Thirdly, the algorithm is also used to classify rainfall levels into four classes as no-rain (0 mm.), light-rain (> 0 - 10 mm.), moderate-rain (>10-35 mm.), and heavy rain (> 35 mm.). Results show that overall classification accuracy of the neuro-fuzzy classifier is satisfactory.**

***Index Terms*— neuro-fuzzy, NEFCLASS, rainfall prediction.**

I. Introduction

Drought is a major problem for people in the central area of Thailand. Normally, it lasts more than three months a year. There is not enough rain for the successful growing of crops or the replenishment of water supplies. Therefore, the Royal Rain Making Project takes responsibility in conducting a number of cloud seeding operations to enhance the precipitation in the areas. However, the success of cloud seeding operations is uncertain as it is necessary to determine or forecast the success rate before any operations are conducted. Several climate factors, precipitation records, and prediction results from cloud models such as the Great Plains Cumulus Model (GPCM), are normally used in making the decision on whether the cloud seeding operation will be launched or not [9]. The event of rainfall occurrence is the most important outcome to evaluate the effectiveness of cloud seeding programs [6]. In this study, the machine learning approaches will be applied for the prediction of rain occurrences. The aims of this study are (1) to present the methodology of a neuro-fuzzy technique to forecast rain occurrence, and (2) identify the relationships between rain occurrences with other weather conditions in the forms of parameters and rules for supporting decisions in seeding operations.

Pramote Luenam is with School of Applied Statistics, National Institute of Development Administration, Bangkok, Thailand; e-mail: pramote.l@ics.nida.ac.th.

Supawadee Ingsriswang is with National Center for Genetic Engineering and Biotechnology, Pathumthani, Thailand.

Lily Ingsrisawang is with Department of Statistics, Faculty of Science, Kasetsart University, Bangkok, Thailand.

Prasert Aungsuratana and Warawut Khantiyanan are with Bureau of the Royal Rainmaking and Agriculture Aviation, Bangkok, Thailand.

II. MATERIALS AND METHODS

A. Data Preparation

This study uses various combinations of data sets collected by the Chalermprakiat Royal Rain Making Research Center at Hua Hin, Prachuap Khiri khan, and Thai Meteorological Department, Thailand during 2004 to 2008. Our three integrated data sets consist of: (1) the so-called Meteor, (2) the GPCM+Meteor, and (3) the GPCM+Meteor+CAPPI. The GPCM stands for Great Plains Cumulus Model and provides us with several climate factors. CAPPI stands for Constant Altitude Plan Position Indicator and is a radar display which gives a horizontal cross-section of data at a constant altitude.

The Meteor dataset contains 1,735 daily records of average rain volumes (AVG) from rain gauges at regional weather stations. The GPCM+Meteor data sets, including the upper air observations, seeding operations and the AVG from rain gauges, consist of 1,148 daily records for the central region. Each GPCM+Meteor record contains 66 variables or features including: temperature, pressure, cloud seeding operations, K-Index, relative humidity at the convective condensation level (RH_AT_THE_CCL), convective temperature (CONVECTIVE_TEMP), average relative humidity at the altitude range of 1,000-5,000 feet (MEAN_RH_1000_5000_FT), average wind speed at the altitude range of 10,000-15,000 feet (MEAN_WIND_10000_15000_FT_KTS), average wind direction (MEAN_WIND_20000_25000_FT_DEG), and average relative humidity at the altitude range of 20,000-25,000 feet (MEAN_RH_20000_25000_FT). The GPCM+Meteor+CAPPI data sets were made by linking the GPCM+Meteor data sets with the average intensity values from the radar observations. The GPCM+Meteor+CAPPI dataset for the central region includes 102 records.

Based on the AVG feature, each record in all data sets was then categorized into (a) rain or no rain events, and (b) rainfall levels (none, light, moderate, and heavy).

B. Feature Selection

In order to find the association among weather conditions and rainfall level classification, a correlated-based feature

ISBN: 978-988-17012-8-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

selection (CFS) was also incorporated to filter out some noisy features and obtain the most important features for rainfall prediction. Applied with correlation, the CFS evaluates the relevance of features with respect to their ability to separate the classes. The selected features from the GPCM+Meteor+CAPPI regional data sets were used to assess the performance of prediction models by the neuro-fuzzy classifier.

C. NEFCLASS

NEFCLASS-J, a JAVA-based soft-computing tool, was used to obtain the classifier from the data sets. NEFCLASS, which stands for NEuro-Fuzzy CLASSification, is a neuro-fuzzy approach to derive fuzzy classification rules from a set of labeled data [3]. NEFCLASS has a capability of using prior knowledge in the form of fuzzy sets and linguistic rules [1]. It uses a supervised learning algorithm based on fuzzy error back-propagation [1], [3].

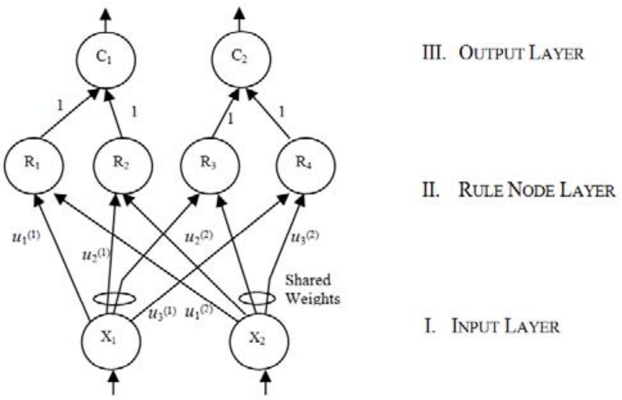

Fig 1. A simple NEFCLASS structure (adapted from [3])

The learning algorithm of NEFCLASS has two stages: (1) rule learning, and (2) parameter learning [8]. In rule learning, a fuzzy rule base is created by using a modified Wang Mendel procedure [7]. All possible antecedents are determined and then an initial rule base is created by finding a proper consequent for each antecedent [2], [7]. Each input feature is partitioned by a given number of initial fuzzy sets [4], [10]. After all input features have been processed; the antecedents are completed by adding fuzzy sets for the symbolical features [4]. In parameter learning, the fuzzy sets are tuned by a simple back-propagation-like algorithm [2]. The algorithm results in shifting the fuzzy sets and in enlarging or reducing their support [4].

A 3-layer structure of NEFCLASS -- with two inputs, four rules, and two outputs -- is illustrated in the Fig. 1. The first layer contains input units. This layer is responsible for outputting external input patterns. The middle layer holds rule nodes. Each node represents a fuzzy rule of the form [4], [5]:

If x_1 is μ_1 and x_2 is μ_2.... and x_n is μ_m then the pattern belongs to class $c_1, c_2, \ldots, c_i$, where $x_1, x_2, \ldots, x_n$ are input variables, and $\mu_1, \mu_2, \ldots, \mu_m$ are fuzzy sets.

The third layer contains output units representing each class [2].

D. Development of the neuro-fuzzy classifier

Two major functions of the neuro-fuzzy model are for predicting rain occurrence and classifying of rainfall levels. The first function is used for predicting rain or no-rain events whereas the second function is used for classifying the rainfall amount into four classes: no-rain (0 mm.), light-rain (> 0 - 10 mm.), moderate-rain (>10-35 mm.), and heavy rain (> 35 mm.). Table I presents the parameter settings used with the classifiers and Table II presents examples of training data used with NEFCLASS in the next-day prediction.

Table I: Parameters for NEFCLASS.

Number and Type of fuzzy sets	3 Bell-shaped for each variable
Aggregation function	Maximum
Size of the rule base	Maximum 100
Rule learning procedure	Best per class
Fuzzy set constraints	Keep relative order Always overlap
Learning rate	0.1
Validation	10-fold cross validation
Stop control	Maximum number of epochs = 500 Minimum number of epochs = 0 Number of epochs after optimum = 100 Admissible classification error = 0

Table II: Examples of training data.

Inputs					Outputs	
MMR_{100}	SH_{idx}	ARH_{10}	MW_{15f}	MR_{22f}	Rain	No Rain
M	L	M	M	H	0	1
L	M	L	L	L	1	0
L	M	L	H	H	0	1

For training, 10-fold cross validation was used. The data set was randomly split into a two-third training set and one-third validation set. The membership functions of input variables are represented by overlapping bell-shaped functions. Each input variable is partitioned into three bell-shaped fuzzy sets: small, medium, and large, where small and large are half bell-shaped while medium is bell-shaped. During the learning procedure without exceeding 500 cycles, the fuzzy sets for the symbolical variables were trained until the error on the validation set could not be further decreased. Fig. 2 provides the comparison of a membership function before and after training.

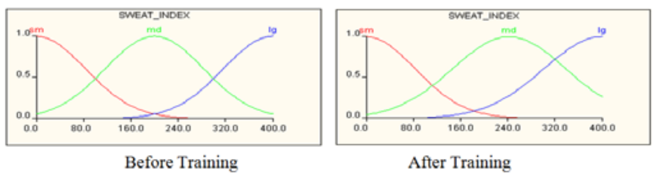

Fig. 2 Comparison of membership functions before and after training.

ISBN: 978-988-17012-8-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Automatic rule learning with best rule per class selection created a rule base with 34, 11, and 5 rules for the prediction model of rain occurrence in same-day, next-day, and next 2-day forecasts, respectively. Resulting from the same learning procedure and parameter settings, the rule base of the rainfall level classification contains 38, 9, and 10 rules in same-day, next-day, and next 2-day forecasts, respectively.

In order to further tune the structure of the rule base in same-day prediction, pruning strategies were applied. The pruning algorithm starts with pruning by correlation and removes variables and rules from the rule base. For the prediction model of rain occurrence, the final rule base contains 10 rules and causes 16 errors on the training set and 4 errors on the validation set. For the classification of rainfall level, the pruning process reduces the number of rules to 15. The final rule base causes 14 errors on the training set and 5 errors on the validation set.

Examples of rules in the final rule base are as follows:

if SWEAT_INDEX is *medium* and
AVG_RH_0_10000_FT is *medium* and
MEAN_WIND_1000_5000_FT_DEG is *medium* and
MEAN_RH_1000_5000_FT is *medium* and
MEAN_RH_5000_10000_FT is *medium* and
MEAN_RH_20000_25000_FT is *medium* then RAIN

if PRECIPITABLE_WATER_TOTAL is *small* and
SWEAT_INDEX is *small*
and AVG_RH_0_10000_FT is *medium* and
MEAN_WIND_1000_5000_FT_DEG is *small* and
MEAN_RH_1000_5000_FT is *medium* and
MEAN_RH_5000_10000_FT is *medium* and
MEAN_RH_20000_25000_FT is *small* then NO RAIN

E. *Model Evaluation*

The classifier is evaluated by its classification accuracy. The C4.5 tree was included as a benchmark to compare the results of the NEFCLASS. Both algorithms were evaluated by their classification accuracy

III. RESULTS

For predicting rain occurrence, the neuro-fuzzy classifier achieves an accuracy of 84.27% - 84.76% in same-day and next-day forecasts, but provides a somewhat higher accuracy of 90.50% in the next 2-day prediction. These results imply that the neuro-fuzzy models perform better or to greater accuracy for the long-term rather range than the short-term and middle-term ranges. Table III illustrates the comparison of the classification accuracy between the NEFCLASS and C4.5 tree. Results show that NEFCLASS, when compared to C4.5 tree, provides more classification accuracy for all three forecasting ranges. Results also indicate that the best accuracy of NEFCLASS (compared to C4.5) on the task of rain prediction is in the long-term forecast.

For the classification of the rainfall level, the neuro-fuzzy classifier, as presented in Table IV, achieves accuracies of 65.82%, 70.24%, and 90.50% in same-day, next-day, and next 2-day forecast, respectively. When compared to C4.5, NEFCLASS provides more classification accuracy across almost all ranges of rainfall forecasting except in the next-day forecast. The results also support the previous findings that NEFCLASS yields the highest classification accuracy in long-term forecasts.

Table III: The overall classification accuracy of the neuro-fuzzy classifier and C4.5 in the prediction of rain occurrence.

	Classification Accuracy		Number of Features
	NEFCLASS	C4.5	
Same-day	84.27	83.33	12
Next-day	84.76	69.23	5
Next 2-day	90.50	75.61	7

Table IV: The overall classification accuracy of the neuro-fuzzy classifier and C4.5 in rainfall level classification.

	Classification Accuracy		Number of Features
	NEFCLASS	C4.5	
Same-day	65.82	61.76	14
Next-day	70.24	72.31	6
Next 2-day	90.50	65.85	10

IV. CONCLUSION

In this paper, we have used the NEFCLASS approach to create a classifier for predicting rain occurrence and classifying rainfall levels. The classifier is very compact and interpretable and its classification accuracy is very good. We have also evaluated and compared the model with the C4.5 tree model. Results show that NEFCLASS forecasting power is more accurate. As evidenced in our results, the methodology can be used to facilitate monitoring of weather conditions and predict rain occurrence over the central region of Thailand, and can be applied to the conduct of appropriate seeding operations in other regions of Thailand.

ACKNOWLEDGMENT

The authors thank the Bureau of Royal Rainmaking and Agricultural Aviation (BRRAA) for providing funds to conduct this study. Also, we would like to thank the Thai Meteorological Department (TMD) as well as many individuals including the director and the administrators of the BRRAA and the director and the officers of the TMD for sharing information and supporting data.

REFERENCES

[1] M. Dixon, Can an integrated ground water vulnerability mapping tool facilitate sensitivity analysis in a spatial domain? Geo Environment. WIT Press, Southampton, UK, 2004.

[2] D. Nauck, and K. Rudolf, NEFCLASS-J :A JAVA-Based Soft Computing Tool. Intelligent Systems and Soft Computing, 2000, pp., 139-160.

[3] D. Nauck, and K. Rudolf, NEFCLASS – a Neuro-fuzzy Approach for the Classification of Data. ACM Symposium on Applied, Nashville, 1995, pp., 461–465.

[4] D. Nauck, U. Nauck, and R. Kruse, NEFCLASS for JAVA - New Learning Algorithms. Seventh International Fuzzy Systems Association, World Congress IFSA '97, Prague, 1997, pp., 50-55.
[5] D. Nauck, F. Klawonn, and R. Kruse, Foundations of Neuro-Fuzzy Systems,Wiley, Chichester, UK, 1997.
[6] L. Ingsrisawang, S. Ingsriswang, P. Luenam, P. Teesaranuwatana, and S. Somchit, The Development of Rainmaking Forecast System over the Central and Eastern River Basin, Thailand. Final Report. Kasetsart University and Bureau of Royal Rainmaking and Agricultural Aviation. Thailand, 2009.
[7] L. X. Wang, and J. Mendel, Generating fuzzy rules by learning from examples. IEEE Trans Syst Man Cybern, 1992, pp., 1414–1427.
[8] S. Serhatlioglu , Z. Bozgeyik , Y. Ozkan , F. Hardalac, I. Guler, Neurofuzzy Classification of the Effect of Diabetes Mellitus on Carotid Artery, Journal of Medical Systems, vol.27(5), October 2003, pp., 457-464.
[9] W. Khantiyanan, Analysis of GPCM forecasting model results. Bureau of Royal Rainmaking and Agricultural Aviation, Thailand, 1996.
[10] M.A. Akcayol, C. Elmas, NEFCLASS-based neuro fuzzy controller for SRM drive, Engineering Applications of Artificial Intelligence, vol.18(5), 2005, pp., 595-602.

ISBN: 978-988-17012-8-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)