Rainfall Prediction in Kemayoran Jakarta Using Hybrid Genetic Algorithm (GA) and Partially Connected Feedforward Neural Network (PCFNN)

Septian Nurcahyo
Telkom School of Computing
Telkom University
Bandung 40257, Indonesia
cahyosn@students.telkomuniversity.ac.id

Fhira Nhita
Telkom School of Computing
Telkom University
Bandung 40257, Indonesia
fhiranhita@telkomuniversity.ac.id

Adiwijaya
Telkom School of Computing
Telkom University
Bandung 40257, Indonesia
adiwijaya@telkomuniversity.ac.id

***Abstract*—The weather changes easily these days, that is difficult to predict. Yet, weather forecasting is the important and useful thing in all aspects of life for instance, in the agriculture field to decide the time of planting. Thus, weather forecast of rain fall intensity particularly in region of Kemayoran Jakarta is conducted in this research. The forecast system built uses Hybrid Genetic Algorithm (GA) and Partially Connected Feedforward Neural Network (PCFNN). In this research, optimum weight and connection neural is optimized with the type of PCFNN. Previously, it has been conducted a research using Evolving Neural Network with all connected neural network and the result was MAPE prediction result rainfall 38.82% or accuracy as 61.18%. Yet, based on the experiment hybrid algorithm between GA and PCFNN, it is gained the MAPE value 35.20% or accuracy as 64.80% with the solution of data missing value in rainfall intensity replaced by 0 (zero) and MAPE 18.48% or accuracy 81.52% for missing value solution in rainfall intensity replaced by mean value.**

***Keywords*—Rainfall; Kemayoran Jakarta; Evolving Neural Network; Genetic Algorithm; Partially Connected Feedforward Neural Network**

I. INTRODUCTION

Previous years, dry and wet seasons in Indonesia, especially in Kemayoran Jakarta could be predicted by only using the calendar, but since the weather system is getting uncertain and can cause disadvantages of many fields, daily activities and many others, thus it is needed a system which can accurately predict the weather changes. In the previous research, it had been building a weather forecast system which used Artificial Neural Network (ANN) with the optimum of Conjugate Gradient, but the result from the research couldn't be used predict the rainfall because of the high Mean Absolute Percentage Error (MAPE) [1,3]. On the other hands, the research on rainfall prediction using algorithm Evolving Neural Network (ENN) which has fully connected feedforward neural network [3], can predict the rainfall with MAPE as 38.82%. in this research, it is hoped that with the implementation of hybrid Genetic Algorithm (GA) and PCFNN which has partially connected feedforward neural network, enables us to produce a better rainfall prediction.

ANN is a part of information process system which has similar characteristics as biology neural network. ANN has node connected one another,and every node connected has its own weight which later on will be updated so that can produce optimum weight to forecast data time series, data which has connection with the previous data. Evolutionary Algorithms (EAs) is a part of Evolutionary Computation (EC), which is optimum algorithms. Evolutionary Algorithms (EAs) mechanism is inspired by biology evolution, such as reproduction, mutation, cross-over, and selection. An Individuals in a population has important role to get solution on optimum matter, and fitness function to determine where the solution "lives".

Hybrid Genetic Algorithm and PCFNN are hybrid from Neural Network Algorithm and Evolutionary Algorithm. Algorithm used in evolutionary Algorithm is Genetic Algorithm, which combine both algorithms which produce optimum conection and weight [4,5].

In this research, Hybrid Genetic Algorithm and PCFNN will produce connection and weight partially connected so that can predict the rainfall in Kemayoran Jakarta for 7 days ahead.

II. DATA AND METHODOLOGY

This system will be built in several stages; raw data, preprocessing, learning using Genetic Algorithm and PCFNN to search the most optimum connection and weight partially connected based on evaluation conducted which produce minimum MAPE [2], the test using connection and partially connected weight which is gained from learning, and rainfall prediction stage using optimum connection and partially connected weight as described in Fig. 1.

978-1-4799-3580-2/14/$31.00 ©2014 IEEE

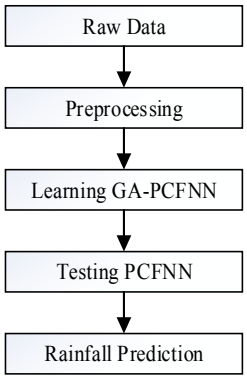

Figure 1. General Scheme of the System

A. Raw Data

Data is used to covers the data from temperature variable (^{0}C), velocity of the average wind (knots), length of sun shines (%), air pressure (milibar), humidity (%), and rainfall intensity (millimeter). The data is gained from Indonesian Agency for Meteorology Climatology and Geophysics (BMKG) region of Kemayoran Jakarta which is available from January 1st 2007 to February 29th 2012.

B. Preprocessing

Missing value can be described as "missing' data or information in the time of collecting data or because of the other factors. To solve the missing value, it can be done by erasing the data which has the missing value within, but this way is effective if only the missing value data is a small part from the whole data. If there is more missing value then this procedure is not possible to do. The other solution for solving this is by filling the missing data with center measure value like mean.

The solution of missing value is by replacing data which has missing value in temperature variable, average wind velocity, length of sun shines, air pressure and humidity will be replaced by mean from each variables yet rainfall intensity will be solved in 2 ways: replacing the data missing with 0 (zero) or value of rainfall variable mean.

After solving the missing data, the weather data of Kemayoran Jakarta will be normalized by changing the range value [0…1] which is aimed to minimize the error. The equation of normalize as follows.

$$Xn = \left(\frac{X_i - MinX}{MaxX - MinX}\right) \quad (1)$$

With:
X_n = normalized data n
X_i = actual data i
MinX = minimum value of X variable
MaxX = maximum value of X variable

C. Learning GA-PCFNN

Learning stage is purposed in producing optimum connection and partially connected weight. In this stage Genetics Algorithm and PCFNN are used. This algorithm results in connection and optimum weight based on the fitness value. The following is learning GA-PCFNN flowchart in Fig.2 :

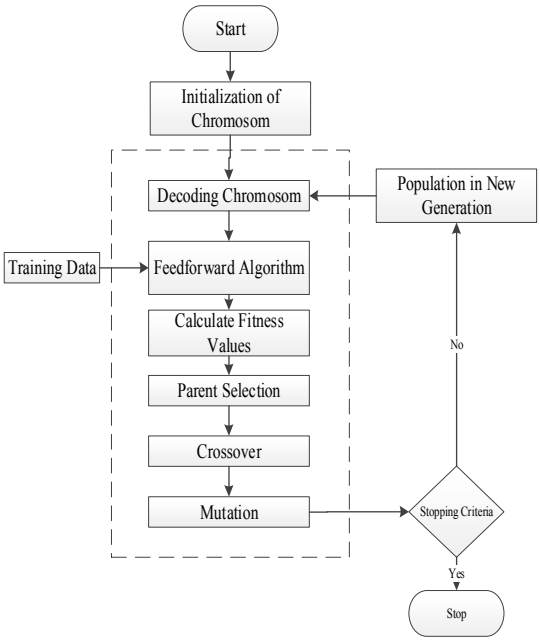

Figure 2. Scheme of GA-PCFNN Learning

The explanation of Fig. 2 as follows :

Initialization of chromosome, where each individuals chromosome in population matrix 1x196 with matrix 1x10 binary represent followed by matrix 1x4 represent real followed by matrix 1x10 represent biner and repeated until formed matrix 1x196, as described in Fig.3.

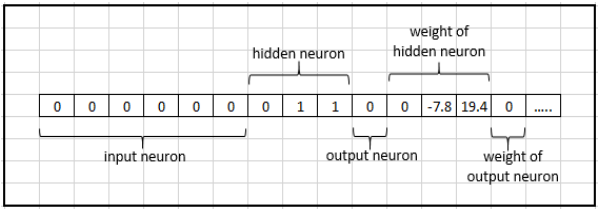

Figure 3. Individuals Representation

Chromosome decode, where will be formed matrix 1x196 from initial population then is decoded using direct-decode into matrix 14x14 by taking matrix 1x14 and putting on the first line then take next matrix 1x14 and put on the second line, this procedure is repeated until formed matrix 14x14 which represents a connection and partially connected weight. The following is the figure of connection and partially connected weight, as described in Fig.4 :

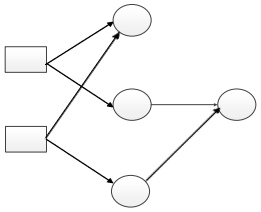

Figure 4. Partially Connected

Connection and partially connected weight then evaluated using algorithm feedforward to get the most minimum error value from all connection and weight being evaluated. Fitness value is searched from each connection and partially connected weight using MAPE value which has already been got from learning process. The selection of parents is aimed to select couple parents with algorithm Roulette-Wheel which focus on probability of fitness value. Cross-over from parents will produce a child with new genetic variation in chromosome yet still has the possibility to have the same gen as the parents with the focus on cross-over probability (Pc) given. Mutation is done to change the gen of 0 to 1 and vice versa for connection and partially connected weight and random change for weight based on mutation probability (Pm). Selection of survivor is the selection where the solution (Individuals) will survive and get in to the next generation. The condition will stop from learning stage if it has fulfilled the maximum Individuals evaluated.

D. Testing GA-PCFNN

After doing the learning stage, it is gained the most optimum connection and partially connected weight which will be used in the testing stage using feed forward, as desribed in Fig.5.

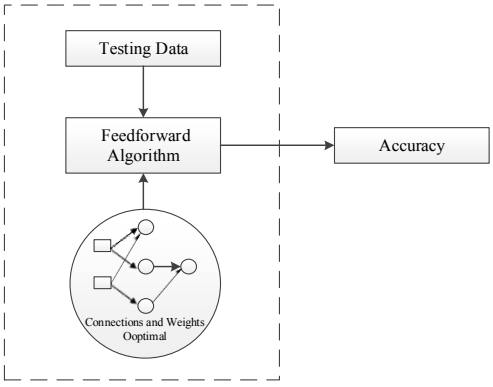

Figure 5. Scheme of Testing GA-PCFNN

The testing stage will result in the accuracy prediction of rainfall in percentage (%) with this following equation (2) :

$$Accuracy = 100 - \left[\frac{1}{n}\left(\sum_{i=1}^{n}\left|\frac{A_i - T_i}{A_i}\right|\right) * 100)\right] \qquad (2)$$

With:

n = the amount of data
i = 1,2, … n
A_i = i'th actual data
T_i = i'th predicted data

E. Rainfall Prediction

The rainfall prediction is almost similar to the testing which uses the most optimal connection and partially connected weight gained from the training process. The data used to predict are temperature, the average of wind speed, sunshine duration, air pressure, air humidity, and previous day rainfall that generates the recent rainfall. The process is conducted for seven consecutive days, as desribed in Fig. 6.

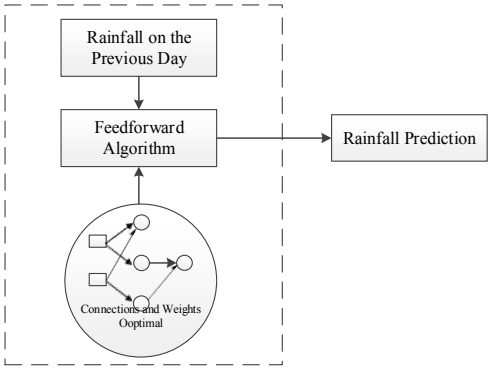

Figure 6. Scheme of Prediction

III. EXPERIMENTS AND RESULTS

In this study, there are several experiment scenarios for probability crossover, probability mutation, number of generation, and number of populations based on table I, table II, and table III below:

TABLE I. SCENARIOS FOR PROBABILITY CROSSOVER (PC) AND MUTATION (PM) COMBINATION

Combination	Experiment								
	1	2	3	4	5	6	7	8	9
Pc	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
Pm	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9

TABLE II. 1ST SCENARIOS FOR NUMBER OF POPULATIONS AND GENERATIONS

Combination	Experiment				
	1	2	3	4	5
Number of Generation	1000	2000	2000	2000	2000
Number of Population	200	160	100	80	50

TABLE III. 2ND SCENARIOS FOR NUMBER OF POPULATIONS AND GENERATIONS

Combination	Experiment				
	1	2	3	4	5
Number of Generation	1000	800	625	500	400
Number of Population	200	250	320	400	500

The following is the explanation on experiment scenarios about the portion of data training-testing and the results for each scenarios.

A. Missing Value at Rainfall Replaced by Zero

- Scenario 1 will use training data up to 50 % of the entire data and the rest will consist of testing data, the maximum number of evaluated Individuals of 200,000, as well as the combination of Pc and Pm as provided in table I.
- Scenario 2 will use training data up to 75 % of the entire data and the rest will consist of testing data, the maximum number of evaluated Individuals of 200,000, as well as the combination of Pc and Pm as provided in table I.

Scenario 1 and scenario 2 will be used in order to obtain the most optimal connection and partially connected weight applied with the best combination of Pc and Pm in table I. After having the best combination of Pc and Pm which generates the optimal connection and partially connected weight, there will be another test using a combination of the number of generations and populations as follows :

- Scenario 3 will use training data up to 50% of the entire data and the rest will consist of testing data, Pc=0.8, Pm=0.2, as well as the maximum number of evaluated Individuals of 200,000, with the combination as provided in table II.
- Scenario 4 will use training data up to 75% of the entire data and the rest will consist of testing data, Pc=0.8, Pm=0.2, as well as the number of maximum evaluated Individuals of 200,000, with the combination as provided in table III.

Scenario 3 and scenario 4 will be used to gain the most optimal connection and partially connected weight applied with the combination of the perfect number of generations and populations from table II and III.

B. Missing Value at Rainfall Replaced by Zero

- Scenario 5 will use training data up to 50% of the entire data and the rest will consist of testing data, the maximum number of evaluated Individuals of 200,000, as well as the combination of Pc and Pm as provided in table I.
- Scenario 6 will use training data up to 75% of the entire data and the rest will consist of testing data, the maximum number of evaluated Individuals of 200,000, as well as the combination of Pc and Pm as provided in table I.

Scenario 5 and scenario 6 will be used to gain the most optimal connection and partially connected weight applied with one of the combination of Pc and Pm as provided in table I. After having the best combination of Pc and Pm which produces the optimal connection and partially connected weight, there will be another test using a combination of the number of generations and populations as follows: :

- Scenario 7 will use training data up to 50% of the entire data and the rest will consist of testing data, Pc=0.8, Pm=0.2, as well as the maximum number of evaluated Individuals of 200,000, with the combination as provided in table II.
- Scenario 8 will use training data up to 75% of the entire data and the rest will consist of testing data, Pc=0.8, Pm=0.2, as well as the maximum number of evaluated Individuals of 200,000, with the combination as provided in table III.

Scenario 7 and scenario 8 will be used to get the most optimal connection and partially connected weight applied with the combination of the perfect number of generations and populations from table II and table III. In this study some experiments are conducted as to predict rainfall. The training of GA-PCFNN generates the optimal connection and weight.

Based on all tested scenarios, there are some connections obtained which can be categorized into 2, namely :

1. Rejected connection

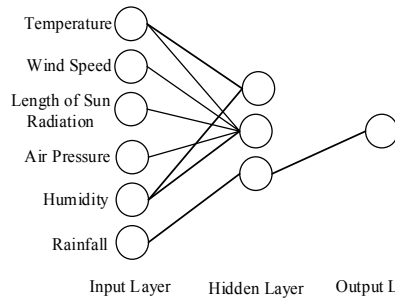

Figure 7. Rejected Connection and Weight Partially Connected

Fig.7 shows one of the examples of connections which only has rainfall parameter passing through hidden layer to output layer while the other parameters are stopped at the hidden layer causing the other parameters to have no influence on rainfall whereas in reality rainfall is influenced by the other parameters.

2. Accepted connection

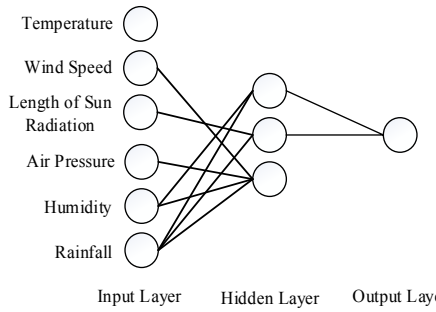

Figure 8. Accepted Connection and Weight Partially Connected

Fig. 8 is considered as one of the examples of accepted connections because besides rainfall, there are also sunshine duration and air humidity passing through the hidden layer and to the output layer. Minimally, there is one other parameter goes through hidden layer to output layer, besides rainfall.

TABLE IV. THE RESULT OF THE EXPERIMENT FOR SCENARIO 1 AND SCENARIO 2

Hybrid GA-PCFNN			
Scenario	MAPE Testing	Pc	Pm
1	31.65%	0.8	0.2
2	40.47%	0.8	0.2

From table IV above, Scenario 1 generates optimal connection and weight with Pc 0.8, Pm 0.2, as well as MAPE testing of 31.65% or testing accuracy of 68.35%. Scenario 2 generates optimal connection and weight with Pc 0.8, Pm 0.2, as well as MAPE testing of 40.47% or testing accuracy of 59.53%.

TABLE V. THE RESULT OF THE EXPERIMENT FOR SCENARIO 3 AND SCENARIO 4

Hybrid GA-PCFNN					
Scenario	MAPE Testing	Pc	Pm	Number of Generation	Number of Population
3	35.67%	0.8	0.2	625	320
4	35.20%	0.8	0.2	800	250

From table V above, Scenario 3 generates optimal connection and weight with Pc 0.8, Pm 0.2, 625 generations, 320 Individuals, as well as MAPE testing of 35.67% or testing accuracy of 64.33%. Scenario 4 generates optimal connection and weight with Pc 0.8, Pm 0.2, 800 generations, 250 Individuals, and MAPE testing of 35.53% or testing accuracy of 64.47%. Based on the result on Table V, the experiments creates connection and partially connected weight with Pc 0.8, Pm 0.2, 800 generations, 250 Individuals, and MAPE testing of 35.53% or testing accuracy of 64.47%. The optimal connection described in Fig. 9.

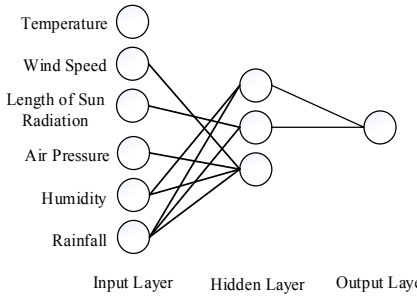

Figure 9. The Optimal Connection from Scenario 1 – Scenario 4

TABLE VI. THE RESULT OF THE EXPERIMENT FOR SCENARIO 5 AND SCENARIO 6

Hybrid GA-PCFNN			
Scenario	MAPE Testing	Pc	Pm
5	14.77%	0.5	0.5
6	12.15%	0.8	0.2

From table VI above, Scenario 5 generates optimal connection and weight with Pc 0.8, Pm 0.2, and MAPE testing of 14.77.65% or testing accuracy of 85.23%. Scenario 6 generates optimal connection and weight with Pc 0.8, Pm 0.2, and MAPE testing of 12.15% or testing accuracy of 87.85%.

TABLE VII. THE RESULT OF THE EXPERIMENT FOR SCENARIO 7 AND SCENARIO 8

Hybrid GA-PCFNN					
Scenario	MAPE Testing	Pc	Pm	Number of Generation	Number of Population
7	28.31%	0.8	0.2	1000	200
8	18.48%	0.8	0.2	1000	200

Scenario 7 generates optimal connection and weight with Pc 0.8, Pm 0.2, 1000 generations, 200 Individuals and MAPE testing of 28.31% or testing accuracy of 71.69%. Scenario 8 generates optimal connection and weight with Pc 0.8, Pm 0.2, 1000 generations, 200 Individuals, and MAPE testing of 18.48% or testing accuracy of 81.52%. Based on the result on table VII, the experiments reach optimal connection and weight with Pc 0.8, Pm 0.2, 1000 generations, 200 Individualss, and MAPE testing of 18.48% or testing accuracy of 81.52%. The optimal connection is as follow in Fig. 10 :

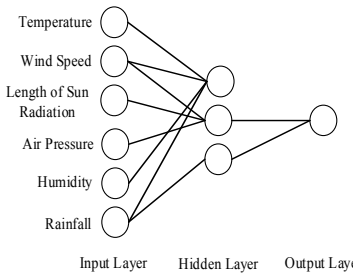

Figure 10. The Optimal Connection from Scenario 5 – Scenario 8

Based on the training of GA-PCFNN in accordance with all testing scenarios, there are various partially connected connections, the connections explain that rain is not only absolutely influenced by all weather parameters such as temperature, wind speed, sunshine duration, air pressure, humidity and rainfall, but rain can also be influenced by several parameters and it is absolutely influenced by the condition of the previous day rainfall.

IV. CONCLUSIONS

In the previous study [3], ENN was able to predict the rainfall with MAPE of 38.82% or the accuracy of 61.18% through all connected networks. Then, in this study, hybrid Genetic Algorithm and feedforward generated connection and partially connected weight with MAPE testing of 35.53% or testing accuracy of 64.47% at which missing value of rain was replaced by 0 (zero) as well as MAPE testing of 18.84% or testing accuracy of 81.52% at which missing value of rainfall was replaced by mean value. It shows that an optimal connection should not be fully connected.

ACKNOWLEDGMENT

The authors would like to thank Telkom School of Computing, Telkom University for financial supporting in this research. The authors would like to thank Indonesian Agency for Meteorology, Climatology and Geophysics (BMKG) for sharing information and supporting data.

REFERENCES

[1] Adiwijaya, U.N. Wisesty, and F. Nhita, "Some Line Search Techniques on the Modified Backpropagation for Forecasting of Weather Data in Indonesia", the Far East Journal of Mathematical Sciences 86:2 pp. 139-148, 2014.

[2] Isik, C., and Kang, S. ,"Partially Connected Feedforward Neural Networks". IEEE Transactions on Neural Networks, Vol.16, No.1, January 2005

[3] F. Nhita, Adiwijaya, and Wisesty U.N, "A Forecasting of Weather Data in Indonesia Using Evolving Neural Network based on Genetic Algorithm", The Second International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE 2014), pp. 78-82, March 2014

[4] Liu, Z., Wang , C., Liu , A., and Niu, Z, "Evolving Neural Network Using Real Coded Genetic Algorithm(GA) for Multispectral Image Classification". *International Conference on Intelligent Computing: Computational intelligence* (pp. 1320–1336), 2004

[5] Rismala, R., Liang, T., and Ardiyanti, A, " Prediction of Malaria Incidence in Banggai Regency", International Conference on Technology, Informatics, Management, Engineering & Environment (TIME-E) p. 94, 2013