A Rainfall Forecasting using Fuzzy System Based on Genetic Algorithm

Fhira Nhita
Computational Science
Telkom Institute of Technology
Bandung 40257, Indonesia
vir@ittelkom.ac.id

Adiwijaya
Computational Science
Telkom Institute of Technology
Bandung 40257, Indonesia
adiwijaya@ittelkom.ac.id

***Abstract*—Weather forecasting information is very crucial in decision making process regarding to activities and works, such as in the field of agriculture to determine initial growing season. Recently, climate change causes trouble in weather forecasting. In this paper, rainfall forecasting system using fuzzy system based on genetic algorithm (GA) is made. The data used within this research is taken from Indonesian Agency for Meteorology, Climatology and Geophysics (BMKG) for Kemayoran area, Jakarta. Those data include temperature, air pressure, rainfall, solar radiation , relative humidity , and wind speed. Based on experiment result, it can be concluded that the combination of GA and Fuzzy for Kemayoran weather data can produce prediction model with higher than 90% accuracy with several population size and crossover probability. This condition is highly affected by input data and conducted classification. Obtained model can be used in predicting rainfall class for the next day that divided into 4 class of no-rain or mild rain (<=20 mm), moderate rain (<=50 mm), heavy rain (<=100 mm), and very heavy rain (>100 mm).**

Keywords—prediction;rainfall;fuzzy system;genetic algorithm

I. INTRODUCTION

Weather forecasting information is crucially needed for decision making in conducting activity or work, for example regarding to argiculture. Due to [3], Indonesia as archipelago country which located in equator area is prone to climate change. Several climate aspect which can be changed are rainfall pattern, sea level, and temperature. Agriculture is most vulnerable to climate change, especially regarding to national food sustainability [3]. Meanwhile current climate changes is hard to be predicted. There are so many research conducted regarding to weather predictionsuch as Bayesian network probability model for weather prediction (Nandar, 2009), The weather prediction method based on Artificial Immune Sytem (Xiang Weiguo, 2010), An Efficient Weather Forecasting System using Artificial Neural Network [2], and A Neuro-fuzzy approach for daily rainfall prediction over the central region of Thailand [6] which classifying rain into 4 classses of no rain (0 mm), light-rain (>0-10mm), moderate-rain (>10-35mm), and heavy-rain(>35mm). Meanwhile in this research Genetic algorithm (GA)-optimized by fuzzy system method is used by categorizing rainfall based on BMKG information. Fuzzy system is used because the weather forecasting information that delivered in electronic and printed media is also obtained from BMKG and contain element of fuzzy, for example today's weather forecast is sunny, cloudy, drizzly, moderate rainy, and highly intensed rainy.

Hybrid Fuzzy and GA are also vastly used in other sector such as researches of A fuzzy genetic algorithm for driver scheduling [4] and Hybrid Fuzzy-Genetic Algorithm Approach for Crew Grouping [5].

In this research, weather forecast for Kemayoran area, Jakarta is assessed which involves weather related data such as temperature, air pressure, rainfall, solar radiation , relative humidity , and wind speed [1,7,9]. In assessing this weather forecast, fuzzy system is used to forecast rainfall for next day such that it produce information, for example class of mild rain has degree of membership value equal to 0.65. In the initial condition, input and output data of this study are not in the form of fuzzy. however, the data has been converted into fuzzy due the data output in the form the rainfall data classification where the shape of fuzzy membership functions is a potential rainfall that occurred.

Fuzzy system is able to do intervension or rationalization for fuzzy or blur data with linguistic variable input such as is able to do intervension or rationalization for fuzzy or blur data with linguistic variable input such as low, medium, and high temprature so the amount of linguistic, type, and function paramenter membership value should be known generally from expert's knowledge. Since this research is not including expert's knoweldge in determining input for fuzzy system so the genetic algorithm will be imposed to obtain optimum input or parameter for fuzzy system [7,8].

The final result of this research is weather forecast information especially rainfall in Kemayoran area for the next day so this information can be used by society for the sake of desicion making in conductiong their activities and works.

II. FUZZY SYSTEM BASED ON GENETIC ALGORITHM

A. Genetic Algorithm

Genetic algorithm can be used as optimization for fuzzy system such as for obtaining the membership value of

linguistic, type and parameter if there is no expert information. Following is the genetic algorithm pseudocode [8] :

```
Population initialization , N chromosom
Evaluate population
Loop until stopping criterion
        Create 1 or 2 copy from the best chromosome
        Loop until stopping criterion
                Choose 2 chromosome
                Crossover
                Mutation
        endloop
endloop
```

B. *Fuzzy System*

Fuzzy system can classifying data which has fuzzy or blur character. Within fuzzy system, initial information which required are amount of linguistic, type and parameter value from membership function which can be obtained from expert or using other learning algorithm. Following are the main stages of FIS (Fuzzy Inference System) :

1) Fuzzification
   Within this process, data in the form of crisp , for example relative humidity = 0.9 will be transformed into linguistic variable of relative humidity = high with the degree of membership = 0.9. GA will be used in this fuzzification process to determine the number of linguistic variable, type and parameter of membership fuction.
   Following are the three membership functions that observed from Matlab :
   - Triangular

$$f(x;a,b,c)=\begin{cases}0, x\le a\\ \dfrac{x-a}{b-a}, a\le x\le b\\ \dfrac{c-x}{c-b}, b\le x\le c\\ 0, c\le x\end{cases} \quad (1)$$

   - Trapezoidal

$$f(x;a,b,c,d)=\begin{cases}0, x\le a\\ \dfrac{x-a}{b-a}, a\le x\le b\\ 1, b\le x\le c\\ \dfrac{d-x}{d-c}, c\le x\le d\\ 0, d\le x\end{cases} \quad (2)$$

   - Phi

$$f(x;a,b,c,d)=\begin{cases}0, x\le a\\ 2\left(\dfrac{x-a}{b-a}\right)^2, a\le x\le \dfrac{a+b}{2}\\ 1-2\left(\dfrac{x-b}{b-a}\right)^2, \dfrac{a+b}{2}\le x\le b\\ 1, b\le x\le c\\ 1-2\left(\dfrac{x-c}{d-c}\right)^2, c\le x\le \dfrac{c+d}{2}\\ 2\left(\dfrac{x-c}{d-c}\right)^2, \dfrac{c+d}{2}\le x\le d\\ 0, x\ge d\end{cases} \quad (3)$$

2) Inference
   in interfence process IF .. Then.. patterned Mamdani method will be used. The rules are also being obtained from GA.
3) Defuzzification, is changing fuzzy value into crisp output.

C. *Hybrid Fuzzy System With GA*

Hybrid aims to overcome the shortcomings of the fuzzy system by using the advantages of the Genetic Algorithm. The advantage of the GA can be used for the learning process that is generating the input to the system such as generate fuzzy membership function and rules that will be used.

III. DATA AND METHODOLOGY

Fig.1 is presenting the steps of this research. System input used is raw weather data in Kemayoran area, Jakarta from the year of 2007 to 2011. The data have 6 weather attribute, which are temperature, air pressure, rainfall, solar radiation and wind speed.

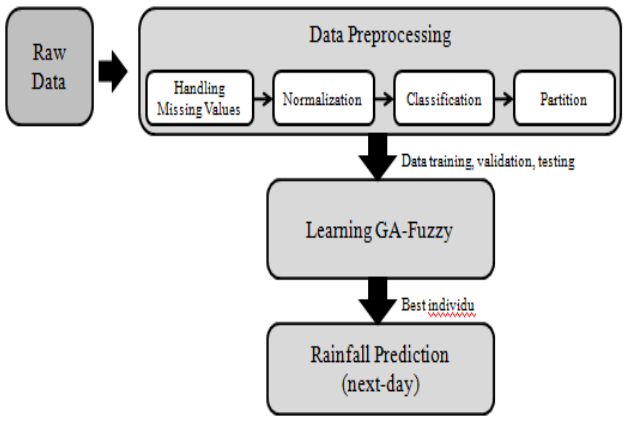

Figure 1. Block Diagram of System

A. Data Preprocessing

Before learning process is conducted, the missing values treatment will be done to raw data. This should be done because existing data has lots of missing values or null values. Following is the result of data audit :

TABLE I. DATA AUDIT

Attribute	% Complete	Valid Records	Null Values
Temperature	100	1826	0
Rainfall	55.91	1021	805
Solar Radiation	96.65	1765	61
Air Pressure	100	1826	0
Relative Humidity	100	1826	0
Wind Speed	99.89	1824	2

Table-1 indicates that the attribute of rainfall, solar radiation, and wind speed has missing or null values so preprocessing should be conducted beforehand to overcome the missing values data by filling "zero" value in every record. After that, the normalization will be done to data using following formula [9] :

$$x' = \frac{0.8(x-a)}{b-a} + 0.1 \tag{4}$$

where:
x = pre normalization data
x' = post normalization data
a = minimum value of overall data
b = maximum value of overall data

Therefore normalized data can be produced within the data range between 0.1-0.9. After that, the next-day rainfall data will be classified based on the information from BMKG as follow:

1) Rainfall <= 0.1mm, categorized as no rain
2) 0.1<Rainfall <= 20mm, categorized as mild rain
3) 20<Rainfall<= 50mm, categorized as moderate rain
4) 50<Rainfall <= 100mm, categorized as heavy rain
5) Rainfall> 100mm, categorized as very heavy rain

The last process of this step is parting data into training, validation, and testing data. Table-II shows the example of training data with the explanation for each attribute as follow: Temperature (TP), Rainfall (RF), Solar radiation (SR), Air pressure (AP), Relative humidity (RH), and Wind speed (WS).

TABLE II. EXAMPLE OF TRAINING DATA

TP	RF	SR	AP	RH	WS	Class (next-day)
0.643	0.227	0.100	0.469	0.224	0.233	1
0.617	0.100	0.535	0.435	0.248	0.300	2
0.657	0.164	0.135	0.491	0.237	0.300	2
0.613	0.122	0.210	0.480	0.244	0.300	2
0.595	0.116	0.100	0.474	0.229	0.433	2
0.621	0.153	0.175	0.503	0.210	0.300	4
0.559	0.333	0.100	0.514	0.219	0.500	5
0.483	0.900	0.100	0.463	0.286	0.367	4
0.475	0.361	0.100	0.463	0.292	0.433	2
0.569	0.115	0.245	0.457	0.222	0.433	2
0.533	0.160	0.100	0.423	0.248	0.300	5
0.581	0.458	0.355	0.463	0.264	0.367	2
0.529	0.102	0.320	0.480	0.244	0.300	3
0.583	0.209	0.175	0.469	0.237	0.367	3

B. Learning GA-Fuzzy

In this step, learning for the training and validation data will be performed. Learning GA is conducted to acquire fuzzy parameters for each attribute which represented within a chromosome with binary representation. Moreover, GA will also be used to obtain fuzzy rules.

C. Rainfall Prediction (Next-day)

The learning result produce best "individual" which contains shape and parameter of membership function, as well as fuzzy rules which will be used to predict the rainfall category based on today's prediction.

IV. EXPERIMENT RESULT

Experiment result is obtained based on examination and scenario parameter which will be explained below. Meanwhile, the data used are taken 2007-2011. Data is divided into three parts, namely training (2007-2009), validation (2010), and testing (2011).

A. Examination Parameter

Test parameters used in learning GA are crossover probability, population size, and maximum individual. In this experiment, several observation have been performed with the Population Size (PopSize) = [50;100], crossover probability (PC)= [0.5;0.7;0.9] and maximum individual = 1000. Meanwhile performance parameter used is fitness function for training data which is comparison of the real class with predicted class, as well as accuracy calculation for testing data.

B. Examination Scenario

Examination scenario is divided 2 scenarios that should be done during data preprocessing which are:

1) Classification of rainfall into 5 different classes of: no-rain (<=0.1 mm), mild rain (<=20mm), moderate

rain (<=50 mm), heavy rain (<=100 mm), and very heavy rain (>100 mm).

2) Classification of rainfall into 4 different classes of: no-rain or mild rain (<=20 mm), moderate rain (<=50 mm), heavy rain (<=100 mm) and very heavy rain (>100 mm). This is performed due missing value for 44% record of rainfall as result of imposing "zero" value so the integration classes is performed.

C. *Examination Result*

The examination result can be seen at table III below.

TABLE III. EXAMINATION RESULT FOR TRAINING DATA

PC	PopSize	Maximum Fitness	
		1st scenario	2nd scenario
0.5	50	65.22	90.46
0.5	100	65.54	90.42
0.7	50	66.68	90.42
0.7	100	66.63	90.60
0.9	50	65.17	90.56
0.9	100	65.35	90.51
Average		65.77	90.49

From the Table-III can be seen that average accuracy of training data in 1st scenario is not satisfying with the average of 65%. This happens probably because the data in rainfall attribute have so many missing value of 44% from total data (look at table-I) where those data has been classified into 1st class (no-rain) in 1st scenario. Whereas, in the second scenario, average accuracy of 90% obtained by integration of classes.

Table-IV shows the example of fuzzy parameter for each attribute from observation PC=0.7 dan PopSize=100 in second scenario. The description of attribute can be look at table-II.

TABLE IV. FUZZY PARAMETER

Attribute	Membership Function	Linguistic Values	Parameters
TP	Trapezoid	LV#1	a=0;b=0; c=0.02362; d=0.08661
		LV#2	a=0.02362;b=0.08661; c=0.5197; d=0.5512
		LV#3	a=0.5197;b=0.5512; c=1;d=1
RF	Phi	LV#1	a=0;b=0; c=0.1102;d=0.7323
		LV#2	a=0.1102;b=0.7323; c=0.8346;d=0.9606
		LV#3	a=0.8346;b=0.9606; c=1; d= 1
SR	Triangular	LV#1	a=0;b=0.991;c=1
		LV#2	a=0;b=0.9921;c=1
		LV#3	a=0.9921;b=1;c=1
AP	Phi	LV#1	a=0;b=0; c=0;d=0.2047
		LV#2	a=0;b=0.2047; c= 0.4724;d=0.4961
		LV#3	a=0.4724;b= 0.4961; c=1;d=1
RH	Phi	LV#1	a=0;b=0; c=0.08661;d= 0.2441
		LV#2	a=0.08661;b=0.2441; c= 0.5276;d= 0.7323
		LV#3	a=0.5276;b= 0.7323; c= 1;d=1
WS	Phi	LV#1	a=0;b=0; c= 0.1496;d= 0.252
		LV#2	a=0.1496;b=0.252; c= 0.2992;d=0.4724
		LV#3	a=0.2992;b= 0.4724; c= 1;d=1

TABLE V. EXAMINATION RESULT FOR TESTING DATA

PC	PopSize	Maximum Fitness	
		Scenario 1	Scenario 2
0.5	50	65.21	95.07
0.5	100	64.93	95.07
0.7	50	64.93	95.34
0.7	100	64.38	95.62
0.9	50	64.11	95.07
0.9	100	65.21	95.07
Rata-rata		64.80	95.21

The experiment result for data testing, at table-V is shown that the accuracy result for second scenario obtain the highest rate of averagely 95%. This occurs because in the the second scenario occurs classes incorporation as explained in the results of training.

V. CONCLUSION

Based on the result of experiment, it is concluded that the combination of GA and Fuzzy System for Kemayoran weather data can produce prediction model with more than 90% accuracy in several different population sizes and crossover probability. This obtained accuracy is highly depending on data which been used as input and classification which been established due to the vast missing value in the data. This result of accuracy shows that hybrid of GA and fuzzy system can perform rainfall forecast with a satisfying result.

ACKNOWLEDGMENT

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

REFERENCES

[1] Adiwijaya, T.A.B. Wirayuda, U.N. Wisesty, Z.K.A. Baizal, "An improvement of backpropagation performance by using Conjugate Gradient on forecasting of air temperature and relative humidity in Indonesia", the Far East Journal of Mathematical Sciences (FJMS), Special Volume 2013 no.1, pp. 57-67

[2] Baboo, Santhosh, "An efficient weather forecasting system using Artificial Neural Network", International Journal of Environmental Science and Development, Vol. 1, No. 4, October 2010 , pp. 321-326

[3] Badan Penelitian and Pengembangan Pertanian Kementrian Pertanian, "Pedoman umum adaptasi perubahan iklim sektor pertanian", ISBN - 978-602-9462-04-3

[4] J. Li, R.S.K. Kwan, "A fuzzy genetic algorithm for driver scheduling", European Journal of Operational Research 147 (2003) pp.334-344

[5] L. Hongbo, X. Zhanguo, A. Ajith, "Hybrid Fuzzy-Genetic Algorithm approach for crew grouping," isda, pp.332-337, 5th International Conference on Intelligent Systems Design and Applications (ISDA'05), 2005

[6] L.Pramote, I.Supawadee, I.Lily, A.Prasert, and K.Warawut, "A Neuro-Fuzzy approach for daily rainfall prediction over the central region of Thailand", Proceedings of the International Multi Conference of Engineers and Computer Scientist 2010 Vol I, IMECS 2010.

[7] Nhita Fhira, "The prediction of Dengue Haemorragic Fever (DHF) in cimahi using hybrid genetic algorithm and fuzzy logic", Jurnal Penelitian and Pengembangan Telekomunikasi June 2011 **16:1** pp. 42-47

[8] Suyanto, "Soft Computing: Membangun Mesin Ber-IQ Tinggi", Informatika Publisher, Bandung, September 2008, ISBN: 978-979-1153-49-2

[9] U.N. Wisesty, Adiwijaya, T.A.B. Wirayuda. "Analisis and implementasi algoritma Conjugate Gradient Polak Ribiere untuk pelatihan backpropagation studi kasus sistem peramalan temperature udara", Jurnal Penelitian and Pengembangan Telekomunikasi December 2010 **15:2** pp.151-155