Model identification for prediction of dengue fever disease spreading using Bat Algorithm and backpropagation

A Damayanti, N L Hidayati and A B Pratiwi

Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia

E-mail: aulid@fst.unair.ac.id

Abstract. Dengue fever disease is one of the public health problems in Indonesia growing rapidly and spreading widely. Dengue fever disease is caused by the dengue virus. The virus is spread by species mosquito Aedes Aegypti and Aedes Alboctipus as primer vector as well as Aedes polynesiensis, Aedes scutellaris and AE (Finlaya) niveus as secondary vector. According to data from WHO, Pacific Asean bore 75 percent from dengue burden in the world during 2004 and 2010, while Indonesia is reported as the biggest second country with dengue fever disease cases between 30 endemic countries. One way to have a better understanding of the problem is by identifying the model based on the known data and do prediction. It is expected that the government would take action based on the prediction result. To solve this problem, the researcher using Bat Algorithm and artificial neural network backpropagation. In order to solve this problem. This paper purposes Bat Algorithm and Artificial Neural Network – Backpropagation to identify the spreading model. The purpose of the identification system using neural network backpropagation is to identify the ODE model of spreading dengue fever disease based on actual data. The first process is the estimation of parameters model using bat algorithm, with inquiring for a numeric solution from the ODE model spread of dengue fever disease as an objective function. The second process is model identification using artificial neural networks and the last, prediction of dengue fever spreading. T Based on the simulation result using dengue fever disease data start from January 2013 until December 2017 the MSE is 0.008 for identification process and 0.1728 for prediction process whereas The MSE value in validation process result is 0.1089 for identification process and 0.1617 for prediction process.

1. Introduction

Disease is a condition of living organisms that are attacked by foreign substances, causing changes in organ function. Disease can also be defined as an abnormal condition of the body which results in the body not being able to work normally [8]. Diseases are classified into two namely, non-infectious diseases and infectious diseases. Infectious disease is very dangerous for human survival because it can kill several residents directly. One example of an infectious disease that is of public concern is dengue fever. According to WHO data in the years 2004-2010, 75 percent of the world's population in the Asia Pacific affected by Dengue Fever. Indonesia is the second-largest country among the 30 countries with the largest endemic areas of Dengue Fever. In 2017 the number of reported cases of Dengue Fever was 68,407 cases with 493 people died. [11]
One effort to minimize the spread of dengue fever is to identify models and predict the spread of dengue fever in the coming year so that it can be used as consideration for future decisions. System identification is the process of identifying the input and output relationships of a system. Prediction is an attempt to predict the situation to come by looking at the situation in the past. There are many ways to identify systems and forecasting, one of which is the Artificial Neural Network (ANN) method. ANN is an information processing method that has certain performance characteristics that are the same as the neural network of living things [1].

Many researchers conclude that ANN is the best method for prediction [7]. Artificial Neural Network (ANN) can also be used in solving nonlinear system identification problems [3]. One method of learning neural networks that can be used for prediction is Backpropagation.

Backpropagation is a network training algorithm that is used to recognize patterns from data during training. Backpropagation is the most popular training algorithm but the computation process is too heavy [5]. Therefore, this research tries to optimize the parameters used in BP training by using the Bat Algorithm in order to obtain a relatively smaller number of iterations.

2. Methodology

The steps used in this study namely determining the DB disease model and data search, parameter estimation using the bat algorithm, identification models using backpropagation neural networks and data validation, and prediction using backpropagation neural networks and data validation.

2.1 Mathematical model and data

The mathematical model of the spread of dengue fever refers to a journal written by Pandey et al in 2013. The model contains three-compartment namely Susceptible (S), Infected (I), and Recovery (R). Dengue fever data was obtained from the East Java Surabaya Indonesia Health Service for the period of January 2013-December 2017. The data consists of three parts, namely, susceptible population, infected population, and recovery population. The model of dengue fever is as follows,

\[ \frac{dS}{dt} = B - \omega \frac{I}{N} S - \mu S \]  
(1)

\[ \frac{dI}{dt} = \omega \frac{I}{N} S - \varepsilon I - \mu I - \theta I \]  
(2)

\[ \frac{dR}{dt} = \varepsilon I - \mu R \]  
(3)

where :

B : birthrate
\(\omega\) : transmission rate of disease between human
\(\mu\) : rate of human death
\(\theta\) : death rate caused by disease

The total population is \(N = S + I + R\). For the development of living things, it is assumed that \(S, I, R, B \geq 0\); and the parameter values are assumed to be positive at intervals (0, 1).

2.2 Parameter estimation

In this study, the estimation of parameter values of the mathematical model dengue fever spread using the bat algorithm. The bat algorithm is an algorithm that is inspired by the behavior and characteristics of the bat. There are three main characters of bats that are used as inspiration for making bat algorithms. First, most bats use echolocation to detect their prey. Second, in looking for prey bat uses frequency, wavelength, and noise. Third, there are many ways to regulate noise [9]. The steps of the parameter values estimation process are in the following:
1. Data
Data dengue fever spread model that used in this research, was taken from East Java Province in the period January 2013 - December 2019.

2. Parameter Estimation
Based on equations (1), (2), and (3), there are five parameters to be estimated namely, \( \delta \) (birth rate), \( \omega \) (transmission rate of transmission of the disease from infected to susceptible), \( \mu \) (natural rate of human death), \( \varepsilon \) (the rate of recovery from Dengue Fever), and \( \theta \) (death rate due to Dengue Fever).

3. Parameters initialization
Parameters initialization consist of, number of bats \([2, \infty]\), pulse rate value \([0,1]\), loudness value \([0,1]\), the value of minimum frequency and maximum frequency \([0, \infty]\), number of iterations \([0, \infty]\).

4. Generate Position and Velocity of Bat
Generate random real numbers \((0,1)\) as the bat position and bat velocity of a number of bat populations.

5. Numerical Solution
The dengue fever spread models are a system of differential equations so the fourth-order runge-kutta method will be used to obtain numerical solutions that approach the exact solutions of a differential equation.

6. Calculate Mean Magnitude of Relative Error (MMRE)
MMRE is a value that is used to measure the difference between the value of the calculation and the true value. The MMRE formula is as follows.

\[
MMRE = \frac{1}{n} \left( \sum_{i=1}^{n} \frac{|y_i - y_i^*|}{y_i} \right)
\]

where:
- \( n \): the amount of
- \( y_i \): actual data
- \( y_i^* \): data from solution numeric

7. Update New Position
The movement of a bat consists of two ways of flying. First is the global step, which is all bats heading towards the best location [11]. The first position update equation is as follows.

\[
f_i = f_{min} + (f_{max} - f_{min})\beta
\]
\[
\dot{v}_i = \dot{v}_i + (x^* - x_i)\dot{f}_i
\]
\[
\dot{x}_i = x_i + \dot{v}_i
\]

where:
- \( f_{min} \): minimum frequency
- \( f_{max} \): maximum frequency
- \( \beta \): random value \([0,1]\)
- \( x^* \): best global position
- \( \dot{v}_i \): velocity in this time
- \( x_i \): position in this time
- \( \dot{x}_i \): next position
- \( \dot{v}_i \): next velocity

The second way is local search, which is a new location for every bat that runs in the local area [11]. The second position update equation is as follows.

\[
x_{new} = x_{old} + \varepsilon < A^{t+1} >
\]
The control of the auto-switch between the first and second mode of flight is obtained by the turning of two parameters, namely, the loudness $A_i$ and the rate of pulse emission $r_i$ [11]. The equation for updating the loudness and the pulse rate are [9]:

$$A_i^{t+1} = \alpha A_i^t$$

$$r_i^{t+1} = r_i^0[1 - (e^{-r_i^0})]$$

where:
- $\varepsilon$: random number [-1, 1]
- $0 < \alpha < 1$
- $\gamma > 0$ are constant
- $r_0$: random number [0,1]

8. Stop conditions
   - If the number of iterations has been fulfilled then the process is complete, if not then return to step 7

2.3 Normalization process

The application of the normalization process is carried out on dengue fever data with the aim of converting the data into intervals [-1,1] [6]. The formula used is as follows,

$$x = \frac{2(x_p - \text{min}(x_p))}{\text{max}(x_p) - \text{min}(x_p)} - 1$$

where,
- $x$: value of data after normalization
- $x_p$: value of data before normalization
- $\text{min}(x_p)$: minimal value of data before normalization
- $\text{max}(x_p)$: maximal value of data before normalization

2.4 Denormalization process

The purpose of the denormalization process is to convert the calculated output data with artificial neural networks into the same value as the original data [6]. The formula used for data denormalization is as follows.

$$x = \frac{(x_p + 1)(\text{max}(x_p) - \text{min}(x_p))}{2} + \text{min}(x_p)$$

where,
- $x$: value of data after denormalization
- $x_p$: value of data before denormalization
- $\text{min}(x_p)$: minimal value of data before normalization
- $\text{max}(x_p)$: maximal value of data before normalization

2.5 Backpropagation Algorithm

Backpropagation consists of the feedforward process and the backpropagation process. The backpropagation neural network architecture is as shown Figure 1.
In the feedforward process, each unit \((x_i)\) will receive an input signal and will spread the signal to the hidden layer \((z_j)\). Each hidden layer that is entered will be calculated the value of activation and will send a signal \((z_j)\) to the output. After that, the incoming output \((y_k)\) will be calculated the activation value \((y_k)\) [1].

During the training process, each output will be searched for the difference between \(y_k\) and \(t_k\) to determine the error value. If the specified error does not match the training error, the weight and bias will continue to be updated until an expected error is obtained. The number of input layers and hidden layers used in this research are two, the activation function used is the bipolar activation function:

\[
y = f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}
\]  

derivative of the activation function:

\[
f'(x) = [1 + f(x)][1 - f(x)]
\]

where:

- \(y\) : output value
- \(x\) : input values

The stages of backpropagation learning:

1. Parameters initialization

Parameters used in backpropagation learning are weights and bias. Initial weights and biases were obtained from the results of the process of discretizing the spread of dengue fever models at intervals \([-1, 1]\). The results of this learning process are in the form of optimal weights and biases, which are used as initial weights and biases to predict the spread of dengue fever.

2. Feedforward

   i. Calculate outputs in hidden unit

      \[
z_{\text{net}_j} = v_{j0} + \sum_{i=1}^{n} x_i v_{ji}
      \]

      \[
z_1 = f \left( z_{\text{net}_j} \right) = \frac{1}{1 + e^{-z_{\text{net}_j}}} 
      \]

   ii. Calculate all outputs in the unit network

      \[
y_{\text{net}_j} = w_{k0} + \sum_{j=1}^{p} z_j w_{kj}
      \]

      \[
y_k = f \left( y_{\text{net}_k} \right) = \frac{1}{1 + e^{-y_{\text{net}_k}}}
      \]

3. Backpropagation error

   i. Calculates the factor \(\delta\) output unit based on the error in each output unit \(y_k\), \(k = 1, ..., m\).

      \[
      \delta_k = (t_k - y_k)f'(y_{\text{net}_k})
      \]

      \[
      = y_k(t_k - y_k)(t_k - y_k)
      \]
ii. Calculates the weight correction that will be used to update the initial weight
\[ \Delta w_{jk} = \alpha \delta_k z_j \] (20)
where
k = 1, 2,..., m
j = 0, 1,..., p

iii. Calculates the bias correction that will be used to update the initial bias value
\[ \Delta w_{0k} = \alpha \delta_k \] (21)
where,
k = 1, 2,..., m

iv. Calculates the hidden unit δ factor based on the error in each hidden unit
\[ \delta_{in,j} = \sum_{k=1}^{m} \delta_k w_{jk} \] (22)
\[ \delta_j = \delta_{in,j} f'(z_{in,j}) \] (23)

v. Calculates the weight correction to be used to update the weight of vij
\[ \Delta v_{ij} = \alpha \delta_j x_i \] (24)

vi. Calculates the correction bias that will be used to update \( v_{0j} \)
\[ \Delta v_{0j} = \alpha \delta_j \] (25)

4. Update Weight and Bias
i. Calculates the change in line weights that goes to the output unit:
\[ w_{k,j}(baru) = w_{k,j}(lama) + \Delta w_{kj} \] (26)
where,
k = 1,2,...,m; j = 0,1,..., p

ii. Calculates the change in line weights that lead to hidden units:
\[ w_{j,i}(baru) = v_{j,i}(lama) + \Delta w_{ji} \] (27)
where,
j = 1,2,...,p; j = 0,1,...,n

iii. Check stopping condition

2.6 Model validation

In this process, a MMRE calculation is done using equation (4), where \( y_i^* \) is the result of backpropagation proses and \( y_i \) is actual data.

The steps to identify the model and forecasting are shown in the following figure
Figure 2. Flowchart of model identification and forecasting

3. Result and Discussions

The data is used in this study are data of dengue fever patients obtained from the East Java Health Office in the period of January 2013-December 2019. The amount of data used is 180 data, the number of healthy or susceptible populations to DB is 60 data, the number of the infected population of DB is 60 data, and the number of recovered populations of DB disease is 60 data.

The data in Table 1 will be carried out system identification and forecasting process. The solution to this problem consists of three main processes namely parameter estimation, system identification, model validation, forecasting, and forecast validation.
Table 1. Data on Spread of dengue fever disease in East Java

| No | Months | Total of Healthy Individuals | Total of Dengue Fever Infected Individuals | Total of Dengue Fever Recovered Individuals |
|----|--------|------------------------------|------------------------------------------|------------------------------------------|
| 1  | Jan-13 | 38698320                    | 3272                                     | 3246                                     |
| 2  | Feb-13 | 38699775                    | 2549                                     | 2514                                     |
| 3  | Mar-13 | 38700723                    | 2067                                     | 2048                                     |
| ...| ...    | ...                         | ...                                      | ...                                      |
| 59 | Nov-17 | 39506389                    | 494                                      | 487                                      |
| 60 | Des-17 | 39506255                    | 562                                      | 553                                      |

In the process of parameters estimation using the bat algorithm to determine the optimal parameter's value that will be used in the identification process. There are five parameters in the mathematical model of the spread of SIR type Dengue fever that is estimated the value. After the optimal parameters are obtained, these parameters are converted into weights and biases in the identification process. After the identification process is complete that is obtained optimal weight and bias, the next process is model validation. The results of the model identification process using the bat algorithm and Backpropagation Neural Network can be seen in figure (3), (4), and (5).

Figure (3) states the simulation results from identifying the model for susceptible populations and actual data. Figure (4) states the simulation results from identifying the model for infected populations and actual data. Figure (5) states the simulation results from identifying the model for recovered populations and actual data. The graph shows that the model can be identified well.

After the identification process is complete, the forecasting process is carried out. Optimal weights and biases from the identification model results are used as initial weights and biases for predicting processes.

In the predicting process, the data is divided into two parts, 70% of data for training and 30% of data for prediction testing. In the training process the initial weights and biases used are weights and biases obtained from the identification process. after training, the prediction test is carried out. The weights and biases used when the prediction test is the best weight and bias from the prediction training results.

Figure 3. Result of model identification of susceptible populations
Simulation results from the prediction test process are shown in figure (6), figure (7), and figure (8).

**Figure 4.** Result of model identification of infected populations

**Figure 5.** Result of model identification of recovery populations

**Figure 6.** Result of forecast of susceptible population
Based on the figure, the smallest MSE values obtained in the identification and prediction process are 0.008 and 0.1728 and the validation errors are 0.1089 and 0.1617, respectively. It means that system can approach the target and recognize actual data pattern. Table 2 shows the results of MSE and validation errors for model identification and prediction with different number of iterations and learning rate.

| no | Iteration | Learning rate | MSE | Validation error |
|---|---|---|---|---|
|   |   |   | I | P | I | P |
| 1 | 10 | 1 | 0.1438 | 1.0636 | 0.2256 | 0.6759 |
| 2 |   | 0.1 | 0.0102 | 0.2043 | 0.1417 | 0.2576 |
| 3 |   | 0.01 | 0.2101 | 0.2838 | 0.5329 | 0.3276 |
| 4 |   | 0.001 | 1.103 | 0.8078 | 1.9869 | 0.7011 |
| 5 | 30 | 1 | 0.3726 | 1.0563 | 0.3341 | 0.3967 |
| 6 |   | 0.1 | 0.0196 | 0.2109 | 0.2173 | 0.2501 |
| 7 |   | 0.01 | 0.0509 | 0.2021 | 0.2109 | 0.209 |
| 8 |   | 0.001 | 0.3784 | 0.962 | 0.2501 | 0.3241 |
| 9 | 500 | 0.1 | 0.0155 | 0.265 | 0.2132 | 0.4344 |
|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 10| 0.01| 0.052| 0.1888| 0.2883| 0.2187|
|11 |0.001| 0.0844| 0.1948| 0.3452| 0.2267|
|12 | 0.1 | 0.0059| 0.2285| 0.097 | 0.3297|
|13 | 0.01| 0.008 | 0.1728| 0.1089| 0.1617|
|14 | 0.001| 0.0191| 0.1746| 0.2026| 0.1974|
|15 | 0.01| 0.0882| 0.2471| 0.1414| 0.1613|
|16 | 0.001| 0.0366| 0.225 | 0.1631| 0.1969|

I: Identification
P: Prediction

4. Conclusion

Based on the results and discussion obtained that the spread of DB disease in East Java can be identified and predicted using the Backpropagation algorithm. The process carried out is parameter estimation, model identification, validation of identification results, prediction, and validation of prediction results. Based on the simulation results, it shows that the system can recognize data patterns with validation errors of 0.1728 and 0.1617 on the 1000th iteration.

References

[1] Fausett L 1994 Fundamentals of Neural Networks (Upper Saddle River: Prentice Hall)
[2] Foss T 2003 A Simulation Study of The Model Evaluation Criterion MMRE IEEE Transactions on Software Engineering 985-995.
[3] Janczak A 2005 Identification of Nonlinear Systems Using Neural Networks and Polynomial Models (New York: Springer)
[4] Pandey A, Mubayi A, and Medlock J 2013 Comparing vector host and SIR Models for Dengue Transmission Math Biosci. 246 (2) 225-259
[5] Prasetyo E 2012 Data Mining Konsep dan Aplikasi Menggunakan Matlab (Yogyakarta: C.V Andi Offset)
[6] Siang J 2005 Jaringan Syaraf Tiruan dan Pemrograman menggunakan Matlab (Yogyakarta: Penerbit Andi)
[7] Sun Z L, Choi T M, Au K F, and Yu Y 2008 Sales Prediction using Extreme Learning Machine with Application in Fashion Retailing Decision Support System 411-419
[8] Timmreck T C 2002 An Introduction to Epidemiology Third Edition (Boston: Jones And Bartlett Publishers)
[9] Yang X S 2010 A New Metaheuristic Bat Inspired Algorithm in Gonzalez J R, Pelta D A, and Cruz C Nature Inspired Cooperative Strategies for Optimization (NICS0 2010) (Berlin: Springer) pp. 65-74
[10] Yang X S 2018 Nature Inspired Algorithms and Applied Optimization (Switzerland: Springer Nature)
[11] Yoeyoen K T W 2017 Infodatin Penyakit Demam Berdarah (Jakarta: Kementerian Kesehatan Republik Indonesia)