Optimizing Adaptive Neuro Fuzzy Inference System (ANFIS) parameters using Cuckoo Search (Case study of world crude oil price estimation)

M Y Anshori¹, D Rahmalia², T Herlambang³, and D F Karya⁴

¹,⁴Department of Management, Universitas Nahdlatul Ulama Surabaya
²Department of Mathematics, Universitas Islam Darul Ulum Lamongan
³Department of Information System, Universitas Nahdlatul Ulama Surabaya

E-mail: ¹yusak.anshori@unusa.ac.id, ²dinitarahmalia@gmail.com, ³teguh@unusa.ac.id

Abstract. There are some methods that have found for estimating data and one of them is Adaptive Neuro Fuzzy Inference System (ANFIS). In estimation using ANFIS, there are some initial parameters such as premise parameters (nonlinear) and consequent parameters (linear) which should be fixed to be trained forward and backward by gradient descent. In this research with case study of world crude oil price estimation, initial ANFIS parameters will be optimized by Cuckoo Search method. Cuckoo Search uses reproduction strategy i.e. laying their eggs in the other bird’s nest. When the eggs are hatched, their chicks are fed by other birds. In Cuckoo Search method, initial ANFIS parameters is represented as bird nest position. Based on simulation, Cuckoo Search method can optimize initial ANFIS parameters giving the best estimation both of training data and testing data in world crude oil price estimation.

1. Introduction

Crude oil is natural resources that have very high demand. Every country is trying to get crude oil since it gives multiply effects in economy and sustainability in the country. There is transmission channels between business cycles and oil prices [15,17]. Most of human living’s aspects relate to energy especially oil and natural gas. Some countries indirectly compete and war among countries to get the oil. Armed conflicts tend to be viewed primarily in geopolitical terms, but they can have profound and long lasting socio-economic consequents [16]. There are many products resulted from crude oil such as liquid petroleum gas (LPG), aviator turbine (avtur) as fuel of aircraft, aviation gasoline (avgas), kerosene, fuel, diesel fuel, asphalt, paraffin and so on. Therefore, estimation of world crude oil price is important to be done.

There are some methods that have found for estimating data. From previous research, some estimation method which have been applied are Kalman Filter applied to blood estimation [5,18] and stock price estimation [10, 19], disease spread estimation [7], weather forecasting [11], H-infinity applied to world crude oil price estimation [12, 20]. In this research, we will use Adaptive Neuro Fuzzy Inference System (ANFIS) optimized by Cuckoo Search. Data used in this research are world crude oil price.

ANFIS is the network using learning algorithm with Takagi-Sugeno fuzzy inference system model. Fuzzy set was introduced by L.A. Zadeh in 1965. Fuzzy sets different with crisp set in which fuzzy set has membership function. There are many application on fuzzy set such as estimation of humidity...
data [9], clustering data in airport [4], clustering agriculture data [8]. In estimation using ANFIS, there are some initial parameters such as premise parameters (nonlinear) and consequent parameters (linear) which should be fixed to be trained forward and backward by gradient descent [1]. In standard ANFIS, initial nonlinear and linear parameters are commonly determined by trial and error so that in this research Cuckoo Search (CS) algorithm will be used to optimized them [3]. In this research with case study of world crude oil price estimation, initial ANFIS parameters will be optimized by CS method. From previous research, CS can be used in optimal control problem [6]. CS uses reproduction strategy i.e. laying their eggs in the other bird’s nest. When the eggs are hatched, their chicks are fed by other birds [13, 14]. In CS method, initial ANFIS parameters is represented as bird nest position. Based on simulation, CS method can optimize initial ANFIS parameters giving the best estimation both of training data and testing data in world crude oil price estimation with fitness function is mean square error (MSE) in testing data. After CS method obtain optimal initial parameters, then the initial parameters will be applied again in ANFIS to be trained by gradient descent and least square estimation (LSE) so that resulting optimal MSE.

2. Adaptive Neuro Fuzzy Inference System (ANFIS)

ANFIS is the network using learning algorithm with Takagi-Sugeno fuzzy inference system model. Figure 1 is the network design of ANFIS [1].

Let two inputs $x$ and $y$, and one output $f$ and there are two rules:

If $x$ is $A_i$ and $y$ is $B_i$ then $f_i = p_i x + q_i y + r_i$

If $x$ is $A_2$ and $y$ is $B_2$ then $f_2 = p_2 x + q_2 y + r_2$

with $A_1, A_2$ and $B_1, B_2$ are membership functions. Parameter $p_1, q_1, r_1$ and $p_2, q_2, r_2$ are consequent parameters.

**Figure 1.** ANFIS structure.

Layer 1. The output in each node is the degree of membership value from the membership functions.

\[
O_{1,i} = \mu_{A_i}(x) \quad i = 1,2 \quad O_{1,i} = \mu_{B_i}(y) \quad i = 3,4
\]
Membership function that is used is generalized bell membership \( \mu(x) = \frac{1}{1 + \left( \frac{x-c}{a} \right)^{2b}} \) or Gaussian membership \( \mu(x) = \exp \left( -\left( \frac{x-c}{a} \right)^{2} \right) \). The parameters \( a,b,c \) are premise parameters.

Layer 2. The output node is the multiplication of signal.

\[
O_{2i} = w_i = \mu_{A_i}(x) \mu_{B_i}(y) \quad i = 1,2
\]  

(2)

Layer 3. The output node is the normalized firing strength.

\[
O_{3i} = \frac{w_i}{\sum_j w_j} \quad i = 1,2
\]  

(3)

Layer 4. The output node is function defined as:

\[
O_{4i} = \frac{w_i f_i}{\sum_j w_j} = \frac{w_i}{\sum_j w_j} (p_i x + q_i y + r_i) \quad i = 1,2
\]  

(4)

Layer 5. This node combines all signals.

\[
O_5 = \sum_i w_i f_i
\]  

(5)

2.1. Hybrid Method

In 1993, Jang developed hybrid method to train premise parameters and consequent parameters [1]. In forward path, signals move from layer 1 to layer 5 and the Least Square Estimation (LSE) calculates consequent parameters. In backward path, the error rates move backward and gradient descent calculates premise parameters.

2.1.1. Optimization on Premise Parameters

Gradient descent optimizes premise parameters \( a,b,c \) as follows:

Supposethe number of training data is \( p \), so that the sum of square error \( E_p \) is as equation (6).

\[
E_p = \sum_{n=1}^{n_{\text{L}}} \left( T_{m,p} - O^L_{m,p} \right)^2
\]  

(6)

with \( T_{m,p} \) is target value of \( m \)-th component of \( p \)-th data and \( O^L_{m,p} \) is output of \( m \)-th component in the layer \( L \) by \( p \)-th data. For \( N \) training data can be computed by \( E = \sum_{p=1}^{N} E_p \)

Compute gradient descent in layer \( L \) in equation (7) and previous layers in equation (8) using chain rule method.

\[
\frac{\partial E_p}{\partial O^L_{i,p}} = -2 \left( T_{i,p} - O^L_{i,p} \right)
\]  

(7)
\[
\frac{\partial E_p}{\partial \Theta_{i,p}} = \sum_{m=1}^{\#(k)} \frac{\partial E_p}{\partial O_{m,p}^{i+1}} \cdot \frac{\partial O_{m,p}^{i+1}}{\partial \Theta_{i,p}}
\]  
(8)

for each \(1 \leq i \leq \#(k)\) with \(1 \leq k \leq L-1\)

If \(\alpha\) is the premise parameter, then the chain rule can be constructed in equation (9)

\[
\frac{\partial E_p}{\partial \alpha} = \sum_{\delta \in S} \frac{\partial E_p}{\partial O^{\delta}} \cdot \frac{\partial O^{\delta}}{\partial \alpha}
\]  
(9)

All errors in \(N\) training data can be computed in equation (10).

\[
\frac{\partial E}{\partial \alpha} = \sum_{p=1}^{N} \frac{\partial E_p}{\partial \alpha}
\]  
(10)

Update premise parameters using equation (11) and equation (12).

\[
\Delta \alpha = -\eta \frac{\partial E}{\partial \alpha}
\]  
(11)

\[
\alpha_{\text{new}} = \alpha_{\text{old}} - \Delta \alpha = \alpha_{\text{old}} - \left( -\eta \frac{\partial E}{\partial \alpha} \right)
\]  
(12)

with \(\eta\) is the learning rate.

2.1.2. Optimization on Consequent Parameters

Suppose linear combinations of output are:

\[
f = \overline{w}_1 f_1 + \overline{w}_2 f_2
\]

\[
f = \overline{w}_1 \left( p_1 x + q_1 y + r_1 \right) + \overline{w}_2 \left( p_2 x + q_2 y + r_2 \right)
\]

\[
f = \left( \overline{w}_1 x \right) p_1 + \left( \overline{w}_1 y \right) q_1 + \left( \overline{w}_2 x \right) p_2 + \left( \overline{w}_2 y \right) q_2 + \left( \overline{w}_2 \right) r_2
\]

For \(N\) training data, they can be modified to become matrix system \(A\theta = Y\):

\[
\begin{bmatrix}
\left( \overline{w}_1 x \right)_1 & \left( \overline{w}_1 y \right)_1 & \left( \overline{w}_1 \right)_1 & \left( \overline{w}_2 x \right)_1 & \left( \overline{w}_2 y \right)_1 & \left( \overline{w}_2 \right)_1 \\
\left( \overline{w}_1 x \right)_2 & \left( \overline{w}_1 y \right)_2 & \left( \overline{w}_1 \right)_2 & \left( \overline{w}_2 x \right)_2 & \left( \overline{w}_2 y \right)_2 & \left( \overline{w}_2 \right)_2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\left( \overline{w}_1 x \right)_N & \left( \overline{w}_1 y \right)_N & \left( \overline{w}_1 \right)_N & \left( \overline{w}_2 x \right)_N & \left( \overline{w}_2 y \right)_N & \left( \overline{w}_2 \right)_N
\end{bmatrix}
\begin{bmatrix}
p_1 \\
q_1 \\
r_1 \\
p_2 \\
q_2 \\
r_2
\end{bmatrix}
= \begin{bmatrix} f_1 \\
f_2 \\
p_2 \\
q_2 \\
r_2
\end{bmatrix}
\]  
(14)

Solution \(\theta = \begin{bmatrix} p_1 & q_1 & r_1 & p_2 & q_2 & r_2 \end{bmatrix}^T\) can be computed by Least Square Estimation (LSE):

\[
\theta = (A^T A)^{-1} A^T Y
\]  
(15)
3. Cuckoo Search

Yang and Deb discovered Cuckoo Search (CS) algorithm in 2010 [13]. CS uses reproduction strategy i.e. laying their eggs in the other bird’s nest and their chicks are fed by other birds [13,14]. They remove otherhost nest’s eggs so that giving more chance for hatching their eggs.

3.1. Behavior of cuckoo birds

Based on behavior of cuckoo birds, the CS algorithm are:

1. Each cuckoo bird lays an egg in randomly selected nest.
2. The best nest with quality eggs will be kept to the next time.
3. Host bird can find an strange egg with \( p \in [0,1] \). The host bird can abandon the nest to build new nest in other location or throw the egg.

3.2. Cuckoo search algorithm applied to ANFIS parameters

In ANFIS method, there are two types of initial parameters which will be optimized such as nonlinear parameter (premise parameters) \( a, b, c \) and linear parameter (consequent parameters) \( p, q, r \). Premise parameters \( a, b, c \) can be trained and optimized using gradient descent. Consequent parameters \( p, q, r \) can be determined by LSE method.

Therefore the representation of decision variable which will be used in CS as host nest is

\[
X = \begin{bmatrix}
a_1 & b_1 & c_1 & \ldots & a_N & b_N & c_N \\
p_1 & q_1 & r_1 & \ldots & p_M & q_M & r_M 
\end{bmatrix}
\]

With \( N \) is the number of membership function required and \( M \) is the number of rules and fitness function is Mean Square Error (MSE) in testing data.

\[
MSE = \frac{1}{\text{datasize}} \sum_{d=1}^{\text{datasize}} \sum_{m=1}^{m} (T_{dk} - Y_{dk})^2
\]

with \( T_{dk} \) is target value and \( Y_{dk} \) is outputs.

In standard ANFIS, initial nonlinear and linear parameters are commonly determined by trial and error so that in this research CS algorithm will be used to optimized them. Based on behavior of cuckoo birds, the CS algorithm for optimizing initial nonlinear and linear parameters in ANFIS can be designed as follows:

Generate initial population of host nests \( x_i, i = 1, 2, ..., maxpop \) randomly. Each of them represents a candidate solution to the optimization problem with objective function \( f(x) \) [9].

For \( t = 1: \text{tmax} \)

1. Calculate global random walk and generate new nest \( x_i^{t+1} \) using Levy Flight.

\[
x_i^{t+1} = x_i^t + \alpha \otimes \text{Levy}(s, \lambda)
\]

Where \( \alpha > 0 \) is the step size scaling factor. The search steps in terms of random randoms \( \text{Levy}(s, \lambda) \) should be drawn from the Levy distribution. In addition, \( \otimes \) denotes the entry wise multiplication.
\[
Levy(s, \lambda) \sim \frac{\lambda \Gamma(\lambda) \sin \left(\frac{\lambda}{2} \pi\right)}{\pi} \frac{1}{s^{\lambda+1}}, \quad s > 0
\]  

(19)

The letter \( \Gamma \) represents the Gamma function \( \Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \). If \( z = k \) is a positive integer, then \( \Gamma(k) = (k-1)! \)

2. Evaluate the fitness of \( x_{i+1} \)

3. Choose a new nest \( j \) randomly from \( \maxpop \) initial nests. If the fitness of \( x_{i+1} \) better than of \( x_j \), replace \( j \) by \( x_{i+1} \)

4. Abandon some of worst nest and build new ones. It depends on discovery probability parameter \( p_d \). Generate random number \( \epsilon \sim U(0,1) \) uniformly distributed. If \( (\epsilon < p_d) \)
   
   - Create new nest using local random walk
     \[
x_{i+1}^j = x_i^j + \alpha \otimes H(p_d - \epsilon) \otimes (x_i^j - x_i^k)
     \]  
     (20)
     
   Where \( \alpha > 0 \), Where \( H(p_d - \epsilon) \) is a Heavyside function.

   - Evaluate the fitness of \( x_{i+1}^j \) and find the best

End

5. Update the best solution

end

4. **The construction of data**

In estimation of world crude oil price, from the dataset we construct the inputs are world crude oil price at time \( t \) as first input \( (x_1) \) and world crude oil price at time \( t+1 \) as second input \( (x_2) \). The output \( (y) \) is world crude oil price at time \( t+2 \) as in figure 2.

![Figure 2. The construction of world crude oil price data for estimation by ANFIS method.](image)

Before applying ANFIS, we normalize the data between 0-1, so that when ANFIS has been done per epoch, we denormalize the data for computing the mean square error (MSE) as objective function. In ANFIS method, there are two types of initial parameters which will be optimized such as nonlinear parameter (premise parameters) \( a, b, c \) and linear parameter (consequent parameters)
$p, q, r$. CS method will be applied for optimizing them with fitness function is MSE in testing data. After CS method obtain optimal initial parameters, then the initial parameters will be applied again in ANFIS to be trained by gradient descent and LSE so that resulting optimal MSE.

5. Simulation results

Data used are world crude oil price dataset (in USD per barrel) during January 2017 until October 2017 with 200 data. From 200 data, they are splitted into training data (80%) and testing data (20%) [2], then we normalize the data so that the data have interval 0-1. ANFIS is used for estimation the data in training data and testing data.

CS method will be applied for optimizing nonlinear parameter (premise parameters) $a, b, c$ and linear parameter (consequent parameters) $p, q, r$. CS parameters used are:

- The number of nests: 5
- Maximum iteration: 50
- Discovery probability parameter $p_a$: 0.5

Figure 3 is optimization process of CS algorithm for optimizing nonlinear parameter (premise parameters) $a, b, c$ and linear parameter (consequent parameters) $p, q, r$ of ANFIS. ANFIS parameters which is used are as follows:

- Maximum epoch: 30
- The number of membership function (gaussmf): 3 (input 1), 3 (input 2)
- The number of rules: 9
- The number of linear parameter (consequent): 27
- The number of nonlinear parameter (premise): 12

First, there are any nests at random position. At the optimization process i.e. minimization of MSE in testing data using Levy flight, abandoning some of worst nest, building new nest, position of best nest is found. Optimal fitness function i.e. MSE in testing data is 0.516352.

![Figure 3. Optimization process of Cuckoo Search.](image)

After CS method obtain optimal initial parameters, then the initial parameters will be applied again in ANFIS to be trained by gradient descent and LSE in training data so that resulting optimal MSE as in Figure 4.
Figure 4. MSE optimization process in training data.

Figure 5 are the initial membership function using gaussian membership function of world crude oil price at time $t$ and time $t+1$ before trained by gradient descent and LSE. There are three membership functions used, i.e. small indicated by blue curve, medium indicated by red curve, and large indicated by green curve. Figure 6 are the optimized membership function using gaussian membership function of world crude oil price at time $t$ and time $t+1$ after trained by gradient descent and LSE. There are three membership functions used, i.e. small indicated by blue curve, medium indicated by red curve, and large indicated by green curve.

(a) (b)

Figure 5. Initial membership function as input (a) world crude oil price at time $t$ (b) world crude oil price at time $t+1$.

(a) (b)

Figure 6. Optimized membership function as input (a) world crude oil price at time $t$ (b) world crude oil price at time $t+1$. 
Figure 7 shows estimation result on training data in world crude oil price at $t+2$. Figure 8 shows estimation result on testing data in world crude oil price at $t+2$. From the simulation, we obtain the prediction with the mean squared error MSE are:

| Data Type     | MSE       |
|---------------|-----------|
| Training data | 0.8386    |
| Testing data  | 0.5164    |

Figure 7. Estimation result in training data.

Figure 8. Estimation result in testing data.

6. Conclusions
In estimation using ANFIS, there are some initial parameters such as premise parameters (nonlinear) and consequent parameters (linear) which should be fixed to be trained forward and backward by gradient descent. In this research with case study of world crude oil price estimation, initial ANFIS parameters will be optimized by Cuckoo Search method. Cuckoo Search uses reproduction strategy i.e. laying their eggs in the other bird’s nest. When the eggs are hatched, their chicks are fed by other birds. In Cuckoo Search method, initial ANFIS parameters is represented as bird nest position. Based on simulation, Cuckoo Search method can optimize initial ANFIS parameters giving the best estimation both of training data and testing data in world crude oil price estimation. The developments of this research are comparison with the other membership functions.

Acknowledgement
This research was supported by LPPM – Nahdlatul Ulama Surabaya of University (UNUSA).

References
[1] Shin J and Jang R 1993 ANFIS : Adaptive Network Based Fuzzy Inference System IEEE Transaction on System, Man and Cybernetics 23 pp. 665-685
[2] Han J, Kamber M and Pei J 2012 Data Mining Concepts and Techniques (USA : Elsevier)
[3] Karaboga D and Kaya E 2017 Training ANFIS by Using the Artificial Bee Colony Algorithm Turkish Journal of Electrical Engineering and Computer Sciences 25 pp. 1669-1679
[4] Rahmalia D and Herlambang T 2018 Application Kohonen Network and Fuzzy C Means for Clustering Airports Based on Frequency of Flight *Kinetik: Game Technology, Information System, Computer Network, Computing* 3 pp. 229-236

[5] Muhith A, Herlambang T, Anshori M Y, Rizzina R and Rahmalia D 2020 Estimation of Whole Blood and Anti-Hemophiliate Factor using Extended Kalman Filter in PMI Surabaya *Journal of Physics: Conference Series* 1538

[6] Herlambang T, Rahmalia D, Nurhadi H, Adzkiya D and Subchan S 2020 Optimization of Linear Quadratic Regulator with Tracking Applied to Autonomous Underwater Vehicle (AUV) Using Cuckoo Search *Nonlinear Dynamics and System Theory* 20 pp. 282

[7] Rahmalia D, Herlambang T, Rohmah A M and Muhith A 2020 Weights Optimization Using Firefly Algorithm on Optimal Control of Zika Disease from Dengue Symptoms by Vaccination *Journal of Physics: Conference Series*

[8] Rohmatullah A, Rahmalia D and Pradana M S 2019 Klasterisasi Data Pertanian di Kabupaten Lamongan Menggunakan Algoritma K-Means dan Fuzzy C Means *Jurnal Ilmiah Teknosains* 5 pp. 86-93

[9] Rahmalia D and Rohmatullah A 2019 Pengaruh Korelasi Data pada Peralaman Kelembaban Udara Menggunakan Adaptive Neuro Fuzzy Inference System (ANFIS) *Applied Technology and Computing Science Journal* 2 pp. 10-24

[10] Karya D F, Katias P, Herlambang T and Rahmalia D 2019 Development of Unscented Kalman Filter Algorithm for Stock Price Estimation *Journal of Physics: Conference Series* 1211

[11] Rahmalia D et al 2020 Comparison between Neural Network (NN) and Adaptive Neuro Fuzzy Inference System (ANFIS) on Sunlight Intensity Prediction Based on Air Temperature and Humidity *Journal of Physics: Conference Series* 1538

[12] Anshori M Y, Herlambang T, Karya D F, Rahmalia D and Inawati P A 2020 H-Infinity for World Crude Oil Price Estimation *Journal of Physics: Conference Series* 1563

[13] Yang X S, Deb S, and Mishra S K 2018 Multi-Species Cuckoo Search Algorithm for Global Optimization *Cognitive Computation* 10 pp. 1085-1095

[14] Yang X S, Chui Z, and Xiao R 2013 *Swarm Intelligent and Bio-Inspired Computation* (London: Elsevier)

[15] Ftiti Z, Guesmi K, Teulon F and Chouachi S 2016 Relationship Between Crude Oil Prices and Economic Growth in Selected OPEC Countries *The Journal of Applied Business Research* pp. 11-21

[16] Estrada M A R, Park D, Tahir M, and Khan A 2020 Simulation of US-Iran War and Its Impact on Global Oil Price Behavior *Borsa Istanbul Review* 20 pp. 1-12

[17] Yoshino N and Taghizadeh-Hesary F 2014 Economic Impacts of Oil Price Fluctuations in Developed and Emerging Economies *IEEJ Energy Journal* 9 pp. 58-75

[18] Muhith A, Herlambang T, Irhamah and Rahmalia D 2020 Estimation of Thrombocyte Concentrate (TC) and Whole Blood (WB) Using Unscented Kalman Filter *International Journal of Advanced Science and Technology* 29 pp. 25-32

[19] Karya D F, Katias P, Herlambang T 2018 Stock Price Estimation Using Ensemble Kalman Filter Square Root Methods *Journal of Physics: Conference Series* 1008

[20] Katias P, Karya D F, Herlambang T, and Khusnah H 2018 Ensemble Kalman Filter for Crude Oil Price Estimation *Journal of Physics: Conference Series* 1211