Water quality prediction of Lake Toba using Extreme Learning Machine (ELM)

Romi Fadillah Rahmat, Eric Suwarno, and Maya Silvi Lydia
Faculty of Computer Science and Information Technology
University of Sumatera Utara
Medan, Indonesia
romi.fadillah@usu.ac.id, 121402071.es@gmail.com, maya2@usu.ac.id

***Abstract*—Currently, the pollution level occurred in Lake Toba varies from lightly polluted to slightly higher level. Thus, there is an increase of the necessity to manage environmental quality of Lake Toba, especially the water quality. Currently, the water quality assessment is done over the laboratory test by obtaining the sample on selected locations of Lake Toba. While Rahmat *et al.* [1] proposed a real-time assessment of water quality in Lake Toba, a method should be implemented to predict the water quality of Lake Toba based on the data collected from the assessment process. In this research, extreme learning machine (ELM) is implemented to predict water quality in Lake Toba. Compared to the backpropagation algorithm, the result shows that the water quality prediction done by using extreme learning machine ...**

***Index Terms*—extreme learning machines (ELM), water quality, Lake Toba, artificial neural networks.**

I. Introduction

According to Haro *et al.* [2], the pollution has been occurred in Lake Toba in North Sumatera province, Indonesia. The level of pollution in Lake Toba varies from lightly polluted to medium level pollution. Residential waste, along with the industrial waste and water hyacinth population on the lake surface, are the main source of the water pollution in Lake Toba.

The water quality assessment is performed by obtaining the sample from several locations around the coastline of Lake Toba. Each samples will be examined by laboratory test to determine the water quality status of Lake Toba. As this assessment method takes more time to be performed, along with the cost of assessment, a method has to be implemented to reduce the time and the cost of assessment process.

II. Theoretical background

A. Extreme learning machine

According to Sun *et al.* [3], extreme learning machines (ELM) refers to a learning method applied in artificial neural networks. The architecture of neural network utilized in extreme learning machine is single hidden layer feedforward neural networks.

Extreme learning machines is proposed by Huang *et al.* [4] to increase the calculation speed of artificial neural networks, by randomizing the hidden layer. They stated that the feedforward neural network utilizes slow gradient based learning, which results in longer computational time. The randomization of the hidden layer results in faster computational speed, along with higher processing result accuracy.

A single hidden layer feed-forward neural network is defined by (1):

$$f_n(x) = \sum_{i=1}^{n} G_i(x, a_i, b_i) * \beta_i, a_i \in R^d, b_i, \beta_i \in R \tag{1}$$

where $G_i(\cdot)$ refers to the activation function calculated in the ith hidden neuron, a_i refers to the input weight received by the ith hidden neuron from input neuron, b_i refers to the bias weight of the hidden neuron, and β_i refers to output weight of the hidden neuron.

For each additional nodes, G_i is defined from the additional node activation function g, as described by (2):

$$G_i(x, a_i, b_i) * \beta_i = g(a_i \times x + b_i) \tag{2}$$

Equation (3) is implemented when the hidden neuron implements RBF as the activation function.

$$G_i(x, a_i, b_i) * \beta_i = g(b_i \parallel x - a_i \parallel) \tag{3}$$

Suppose a training data set $N = \{(x_i, t_i) \mid x_i \in R_n, t_i \in R_m, i = 1, ..., L\}$, with x_i represents the training data, t_i represents the class label of the sample for each instance, and L is defined as the number of hidden nodes. When implementing the extreme learning machine for training the neural network, the steps are done as follows:

- Assign input weights w_i and biases b_i, where $i = 1, ..., L$,
- Calculate the hidden layer output matrix as H,
- Calculate the output weight, as defined in (4):

$$\beta = H^{\dagger}T \tag{4}$$

  where $T = [t_1, ..., t_N]^T$ and $H^{\dagger}$ refers to the Moore-Penrose inverse of matrix H.

B. Water quality index

III. Methodology

This section describes the methodology of this research. The general architecture is shown by Figure 1.

The steps performed in this research are described as follows:

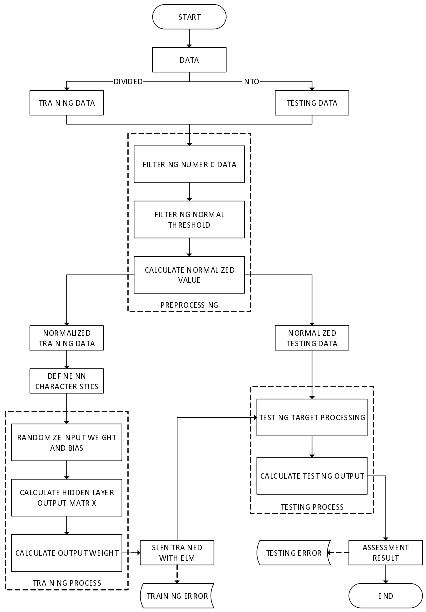

Fig. 1. General architecture of the research

1) Data input: The data utilized in this research is obtained from the research done by Rahmat et al. [1], with the format shown by Figure 2. Each data file will be split into two datasets, namely training dataset and testing dataset, with the ratio of 60:40.

```
1.31,4.58,30.94,34,34,6—80?6.82,
0.31,1.1,30.94,50,34,6.81,
0.58,2.02,30.94,50,34,6.82,
0.84,2.93,30.94,51,34,6.8,
1.1,3.85,30.94,50,34,6.81,
1.36,4.77,30.94,50,34,6.8,
1.62,5.69,30.94,50,34,6.79,
1.89,6.6,30.94,50,34,6.83,
1.99,6.97,30.94,51,34,6.83,
1.99,6.97,30.94,50,34,6.81,
1.99,6.97,30.94,50,34,6.82,
1.99,6.97,31,50,34,6.83,
1.99,6.97,31,47,34,6.83,
1.99,6.97,31,50,34,6.81,
0.18,0.64,85,51,32,*??,
0.73,2.56,27.87,38,33,?R??5.99,
1.04,3.63,27.87,41,33,0,
1.34,4.7,27.94,41,33,0,
1.65,5.76,27.87,41,33,0,
1.95,6.82,27.87,41,33,0.0?,
2.25,7.89,27.87,41,33,0,
2.31,8.1,27.87,41,33,0,
2.31,8.09,27.81,41,33,0,
2.31,8.08,27.81,41,33,0,
2.31,8.08,27.81,41,33,0,
2.31,8.07,27.81,41,33,0,
2.31,8.07,27.81,41,33,0,
2.3,8.07,27.81,40,33,0,
2.3,8.06,27.81,41,33,0,
2.3,8.06,27.81,40,33,0,
2.3,8.05,27.81,40,33,0,
2.3,8.04,27.75,40,34,0,
2.3,8.04,27.75,40,34,0,
2.29,8.02,27.75,40,34,0,
2.29,8.01,27.75,41,34,0,
2.28,7.99,27.75,40,34,0,
2.28,7.97,27.75,40,34,0,
```

Fig. 2. Initial structure of the data

2) Preprocessing: Each training and testing dataset will be preprocessed in order to enable the data to be processed by extreme learning machine. The process is performed in three steps, described as follows:
   - Filtering each row to ensure that each dataset contains fully numeric value;
   - Filtering each row to ensure that contains the normal value of measurement; and
   - Calculating the normalized value of each parameter, which is done by using Eq. 5:

$$A' = \frac{A - A_{min}}{A_{max} - A_{min}} * (D - C) + C \qquad (5)$$

IV. Experiment and Result

V. Conclusion

References

[1] R. F. Rahmat, Athmanathan, M. F. Syahputra dan M. S. Lydia, "Real Time Monitoring System for Water Pollution in Lake Toba," in *Proceedings of ICIC 2016*, Lombok, 2016.

[2] D. D. Haro, Y. Djayus and Z. Harahap, "Kondisi Kualitas Air Danau Toba di Kecamatan Haranggaol Horison Kabupaten Simalungun Sumatera Utara", *AQUACOASTMARINE*, vol. 1, no. 1, 2013.

[3] M. Gao, W. Xu, H. Fu, M. Wang and X. Liang, "A Novel Forecasting Method for Large-Scale Sales Prediction Using Extreme Learning Machine, " *2014 Seventh International Joint Conference on Computational Sciences and Optimization*, Beijing, 2014, pp. 602–606.

[4] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, Dec. 2006.