Levenberg-Marquardt Neural Network for Eye States Detection Based on Electroencephalography Data

Untari N. Wisesty

School of Computing, Telkom University
Bandung, Indonesia
untarinw@telkomuniversity.ac.id

Abstract

The eye state detection is one of various task toward Brain Computer Interface system. The eye state can be read in brain signals. The process of reading the brain signal can be done by several methods, such as Electro Encephalo Graphic (EEG), Electro cortio Graphic (EcoG), and Local Field Potentials (LFPs). From some of the methods, based on several researches, EEG is the most practical and safest to record the brain signals. Therefore, in this paper use EEG Eye State dataset from UCI Machine Learning Repository Database. Dataset is consisting of continuous 14 EEG measurements in 117 seconds. The eye states were marked as “1” or “0”. The mark “1” indicates the eye-closed and “0” the eye-open state. The proposed schemes use Multi-Layer Neural Network with Levenberg Marquardt optimization learning algorithm, as classification method. Levenberg Marquardt method used to optimize the learning algorithm of neural network, because the standard algorithm has a weak convergence rate. It is need many iterations to have minimum error. Based on the analysis towards the experiment on the EEG dataset, it can be conclude that the proposed scheme can be implemented to detect the Eye State. The best accuracy gained from combination variable sigmoid function, data normalization and number of neurons are 31 (95.71%) for one hidden layer, and 98.912% for two hidden layers with number of neurons are 39 and 47 neurons and linear function.

Keywords: Eye State, EEG, Neural Network, Levenberg-Marquardt, optimization.

I. INTRODUCTION

BRAIN Computer Interface (BCI) is a system that using brain signal as a communication and controlling process between brain and computer, and not related to human motoric system (Anderson, n.d) (Alzoubi, 2006). Therefore, BCI are implemented in many aspects to help people that have motoric disability, such as Amyotrophic Lateral Sclerosis (ALS) patients, severe stroke patients that absolutely cannot move, and many diseases that cause the patient cannot use their motoric system. The patients will experience the difficulties to express their will because they cannot speak, write, or communicate with body language. Hence, this has become one of the motivations of researchers to develop and improve the ability of BCI.

The eye state is one of various tasks toward BCI system. The eye state can be read in brain signals. The process of reading the brain signal can be done by several methods, such as Electro Encephalo Graphic (EEG), Electro cortio Graphic (EcoG), and Local Field Potentials (LFPs) (Anderson, n.d). From some of the methods, based on several researches, EEG is the most practical and safest to record the brain signals.

Several papers investigated the difference between the two eye states, whether eyes are open or closed. It was coming to the conclusion that the power in the eye closed state was much higher than that in the eye open state.
Brain Computer Interface (BCI) is a system which allows a person to control special computer applications (e.g. a computer cursor or robotic limb) by using his/her thoughts (Alzoubi, 2006). The idea is to provide a new communication channel to people who are paralyzed, but are cognitively intact, e.g. people suffering from the so called lock-in syndrome. BCIs have been a very active area of research. The research is based on recording and analyzing EEG brain activity and recognizing EEG patterns associated with mental states. For example, imagining a movement of the right hand is associated with a pattern of EEG activity in the left side of the motor cortex. Other frequently used mental tasks are the movement of the left hand, movement of the toes and movement of the tongue. Mental tasks are, carefully chosen so that they activate different parts of the brain, which makes them easier to detect.

The features representing EEG patterns can be obtained using either frequency or nor the frequency domain information of EEG signals (Jue Wang, 2008). Frequency domain information is more widely used in BCI system, for example, complex mental tasks by using asymmetry ratios and power values at four frequency bands: delta, theta, alpha, and beta (Jue Wang, 2008), (Suratgar, 2007). Current methods extract energy information mainly from the whole EEG segments of specific length. However, while mental tasks were performed, EEG segments of any length started from a specific time point are related to different kinds of brain activity information. A convenient way uses a phase space of different dimensions to represent the mental tasks. Figure 1 shows the general BCI system. Electrophysiological signals reflecting brain activity are required from the scalp, from within the brain and processed to measure specific signal features (such as amplitudes of evoked potentials or EEG rhythms or firing rates of single neurons) that reflect the user’s intent. These features are translated into commands that operate a device, such as a word-processing program, a wheelchair, or a neuroprosthesis (Alzoubi, 2006).
B. Eye State Dataset

Dataset acquisitions were conducted in a quiet room. During the experiment, the face of participants was recorded. The experimental procedure was specified as follows (Rosler, 2014):

1. After placing the electrodes on the scalp, the participants were told to sit relaxed, face the camera and change the eye state at free will after clicking the start button.
2. The task was repeated one to two times after a resting period of one minute.

The individual eye state intervals should vary in length and the duration of both eye states should be about the same when accumulated over the entire session.

The duration of the measurement was 140 seconds. Most of the eye states were automatically annotated during the measurement by the video recording program. Only frames which the program could not classify due to bad lighting conditions were later annotated by hand. Both, open or partially open eyes were categorized as open, only completely closed eyes were categorized as close.

The corpus consists of 14,977 instances with 15 attributes each (14 attributes representing the values of the electrodes and the eye state) (Rosler, 2013). The dataset is stored in the corpus in chronological order to be able
to analyze temporal dependencies. There is an obvious difference in amplitude of certain sensors when comparing the range of values for different eye state. For the sensors F7, F3, O2, P8, T8, FC6, and F4, the maximum values of the eye open state are higher than the maximum values of the eye closed state while the minimum values are nearly the same. On the other hand, for the sensors AF3, FC5, T7, P7, O1, F8, and AF4, the minimum values of the eye open state are lower than for the eye closed state while the maximum values are almost same. All sensors have in common that open eye state comes along with a higher value range than the eye closed state while the mean value almost same. As motivated above, sensors could be split into two groups. In the first group, the maximum increases when eyes open. While, in the other group, the minimum decrease in the same event. Most sensors of the first group happen to be located in the right hemisphere while the most of the second group is in the left hemisphere of the brain.

![Fig. 3. Overview of the sensors position and the corresponding behavior group. Blue corresponds to a maximum increase and the red to a minimum decrease when opening eyes.](image)

C. Previous Work

Eye state detection based on EEG data has been developed in the last few years. The first step, (Rosler, 2013) recorded a corpus containing the activation strength of the fourteen electrodes of a commercial EEG headset as well as the manually annotated eye state corresponding to the recorded data. The corpus was tested with 42 different machine learning algorithm using Weka Toolkit and default tuning parameters on their performance to predict the eye state. The best performing classifier, KStar, produces a classification error rate of only 2.7% which is a 94% relative reduction over the majority vote of 44.9% classification error. Standard classifiers such as Naïve Bayes, SMO, logistic regression, or ANNs with a proven track of high classification performance produced rather poor results on this task (over 30% classification error).
Wang et al. (Ting Wang, 2014) applied a promising technique that uses incremental attribute learning (IAL) based on neural network. IAL is a novel machine learning strategy which gradually imports and trains features one by one. IAL exhibited better classification performance in terms of classification error rates in comparison with conventional and some other approaches. The error rate obtained in the final classification result is 27%. Jain et al. (Neha Jain, 2015), performed three algorithms, Linear Support Vector Machine (SVM), SVM polynomial, and SVM RBF. But they did obtained poor performance. Sabanci et al. (Sabanci, 2015), compared two classification method, e.g. k-Nearest Neighbors algorithm and multilayer perceptron neural network models. The classification success rates were calculated for various number of neurons in the hidden layer of a multilayer perceptron neural network model. The highest classification success rate has been obtained when the number of neurons in the hidden layer was equal to 7, and it was 56.45%. The highest classification success rates were calculated with k-nearest neighbors algorithm for different neighborhood values. In the kNN models, the success rate for 3 nearest neighbors were calculated as 84.05%.
III. RESEARCH METHOD

Fig. 5 shows the general system design for eye state detection that implemented in this paper. The system consists of four steps, i.e. input EEG data, preprocessing, classification method (Multi-Layer Neural Network using Backpropagation learning algorithm and optimized by Levenberg-Marquardt optimization) and output system.

![Diagram of the general system design for eye state detection](image)

**A. Dataset Description**

The dataset that used in this research was from EEG Eye State dataset (Rosler, 2013) from UCI Machine Learning Repository Database. Dataset is consisting of continuous 14 EEG measurements in 117 seconds. The eye states were marked as “1” or “0”. The mark “1” indicates the eye-closed and “0” the eye-open state. The duration of the measurement was 117 seconds and the measurement was carried out in a quiet room. During the measurement, the proband was being videotaped. The eye state was manually annotated by analyzing the video recordings aligned with the EEG data.

The dataset consists of 14,977 instances with 15 attributes each (14 attributes representing the value of the electrodes and 1 attribute representing the class of eye state). The instances are stored in the corpus in chronological order to be able to analyze temporal dependencies. 55.12% of the corpus corresponds to the eye open and 44.88% to the eye closed state. Table I shows the value range of the 14 sensors in the corpus.

**TABLE I**

<table>
<thead>
<tr>
<th>Eye State</th>
<th>Closed</th>
<th>Open</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>AF3</td>
<td>4198</td>
<td>4445</td>
</tr>
<tr>
<td>F7</td>
<td>3905</td>
<td>4138</td>
</tr>
<tr>
<td>F3</td>
<td>4212</td>
<td>4367</td>
</tr>
<tr>
<td>FC5</td>
<td>4058</td>
<td>4214</td>
</tr>
<tr>
<td>T7</td>
<td>4309</td>
<td>4435</td>
</tr>
<tr>
<td>P7</td>
<td>4574</td>
<td>4695</td>
</tr>
<tr>
<td>O1</td>
<td>4567</td>
<td>4695</td>
</tr>
</tbody>
</table>

[Fig. 5. General system design Eye States Detection]
B. Preprocessing

The EEG dataset have different values for each sensor. Therefore, they need to normalize into [0, 1]. This step will be the one of testing scenario, that it will be analyzed the performance. Here is the equation to normalize the input data (Wisesty, 2012):

\[
x' = \frac{0.8 (x - a)}{(b - a)} + 0.1
\]

(1)

Where:
- \( x \): value of the data
- \( a \): minimum value
- \( b \): maximum value

C. Multi-Layer Neural Network – Levenberg Marquardt

Artificial Neural Network is one of artificial method that have characteristic similar to biological neural networks. There are three important factors in Neural Network, i.e.:

a. Network architecture
b. Learning algorithm
c. Activation function

Here is the Neural Networks method that used in this paper:

1) Network Architecture

The network architectures that used in this paper is Multi-Layer neural network. This network has at least one hidden layer. In this system uses 14 input neurons, because the dataset has 14 attributes, 1 and 2 hidden layers consist of combination 1 until 50 neurons, and 1 output neuron.

<table>
<thead>
<tr>
<th></th>
<th>P8</th>
<th>T8</th>
<th>FC6</th>
<th>F4</th>
<th>F8</th>
<th>AF4</th>
</tr>
</thead>
<tbody>
<tr>
<td>4147</td>
<td>4174</td>
<td>4130</td>
<td>4225</td>
<td>4510</td>
<td>4264</td>
<td></td>
</tr>
<tr>
<td>4287</td>
<td>4323</td>
<td>4319</td>
<td>4368</td>
<td>4811</td>
<td>4552</td>
<td></td>
</tr>
<tr>
<td>4152</td>
<td>4152</td>
<td>4100</td>
<td>4201</td>
<td>86</td>
<td>1366</td>
<td></td>
</tr>
<tr>
<td>4586</td>
<td>6674</td>
<td>5170</td>
<td>7002</td>
<td>4833</td>
<td>4573</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6. Multilayer Perceptron Architecture.
The connections between neurons are called weights. The optimal values of weights will be evaluated using a learning algorithm. Here is the representation of weights used in this research:

- Weights representation between input layer and hidden layer:

\[
W_{\text{Input}} = \begin{bmatrix}
W_{ih_{1,1}} & \cdots & W_{ih_{1,n}} \\
\vdots & \ddots & \vdots \\
W_{ih_{14,1}} & \cdots & W_{ih_{14,n}}
\end{bmatrix}
\] (2)

- Weights representation between hidden layers:

\[
W_{\text{Hidden}} = \begin{bmatrix}
W_{hh_{1,1}} & \cdots & W_{hh_{1,m}} \\
\vdots & \ddots & \vdots \\
W_{hh_{n,1}} & \cdots & W_{hh_{n,m}}
\end{bmatrix}
\] (3)

- Weights representation between hidden layer and output layer:

\[
W_{\text{Output}} = \begin{bmatrix}
W_{ho_1} \\
\vdots \\
W_{ho_m}
\end{bmatrix}
\] (4)

Where i, j ∈ {1, 2, .. , n}

2) Learning Algorithm

Backpropagation is one of the learning algorithms for Feed Forward Neural Network. The standard backpropagation has three phases:

a. Phase I: Forward propagation
   
   Calculate the output for each neuron from the input layer until the output layer using the specified activation function.

b. Phase II: Backward propagation
   
   The error between the output network and data target denotes the error of the network. The error was backward propagated, starting from the weights in the output layer to the input layer.

c. Phase III: Weight Modification
   
   Update the weights to decrease the network's error.

The standard backpropagation has a weak convergence rate. It is needed many iterations to achieve a minimum error. Hence, backpropagation is usually modified with other optimization methods. One of them is Levenberg-Marquardt optimization (Suratgar, 2007):

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Initialize the weights and parameter μ (learning rate, μ = 0.01 is appropriate)</td>
</tr>
<tr>
<td>2.</td>
<td>Compute the sum of the squared errors over all input F(w).</td>
</tr>
</tbody>
</table>
| 3.   | Solve (2) to obtain the increment of weights Δw. 
      \[ Δw = [J^T J + μI]^{-1} J^T e \] (5) 
      
      Where:
      
      J: the Jacobian matrix
      μ: the learning rate |
| 4.   | Recompute the sum of squared errors F(w) 
      
      Using w + Δw as trial w, and judge: 
      IF trial F(w) < F(w) in step 2 THEN 
      \[ w = w + Δw \] (6) 
      \[ μ = μ \ast β \] (7) 
      Go back to step 2. 
      ELSE 
      \[ μ = \frac{μ}{β} \] (8) |


Where: $\beta$ is decay rate ($\beta = 0.1$ ($0<\beta<1$))
$\mu$ is learning rate.
$F(w)$ is the sum of the squared errors over all input
$\Delta w$ is the increment of weights
$J$ is the Jacobian matrix

3) **Activation Function**

Activation function used to determine output neuron. There are several types of activation functions which have various forms of transformation. But, after analyzing the characteristic of target data, only two activation functions used in this paper, i.e.:

a. Linear function, the input to this function will not transfer to other range data.

$$\mu(v) = v$$  \hspace{1cm} (9)

![Fig. 7. Linear function](image)

b. Sigmoid function, the input to this function will be transferred to range $[0, 1]$.

$$\mu(v) = \frac{1}{1 + e^{-\alpha v}}$$  \hspace{1cm} (10)

Where: $v$ is input value to the activation function.
$\alpha$ is a constant parameter.

![Fig. 8. Sigmoid Function](image)

D. **Output System**

The output value which produced from neural network still in real values. The range values of output system are $[0, 1]$ for sigmoid function and $(-n, n)$ for linear function. Therefore, to obtain a value corresponding to target data (the mark “0” for eye-open state and “1” for eye-closed state), it needs another process called post processing. The post processing used in this paper is:
\[ f(x) = \begin{cases} 0, & x < 0.5 \\ 1, & x \geq 0.5 \end{cases} \] (11)

E. Performance Measurement

In this paper used cross validation method to divide training and testing data. The performance of system was measured using accuracy, because the distribution data between classes was balance. Here is the equation of accuracy:

\[
\text{Accuracy} = \frac{\sum \text{True positive} + \sum \text{True negative}}{\text{Total Data}} \%
\] (12)

IV. RESULTS AND DISCUSSION

This experiment is conducted to analyze the performance of data normalization and Neural Network method that used in this research. Data normalization uses to analyze the influence of equality range of input data, because the dataset used in this research have different range values between attributes. Activation function used to determine output neuron. After analyzing the characteristic of target data, the activation function used is linear and sigmoid function. And the last one of Neural Network parameters is number of hidden layers and number of neurons in hidden layer. It will analyzed the value of two parameters using empirical methods to obtain the best performance. Hence, in this experiment, the variables that used to obtain the best performance are:

- The influence of data normalization (equation 1)
- Activation function in output layer: linear (equation 9) and sigmoid function (equation 10).
- Number of neurons in hidden layer: 1, 2, 3, ..., 50 neurons.
- Number of hidden layers: 1 and 2 hidden layer.

From some variables above, each combination will be observed 5 times and will be taken the best accuracy.

A. The Influence of Data Normalization

![Fig. 9. Influence of Data Normalization to The System Accuracy, the axis x denote activation function and axis y denote the accuracy (%)](image)

Fig. 9 shows the comparison accuracy of data normalization. Based on Fig. 9 indicate data with normalization have better accuracy than raw data. The difference of accuracy is 1.08% for sigmoid function and 0.29% for linear function. This is happen because with normalization the range of input data becomes equal to other, so that the classification method was easier to classify the data.
B. The Influence of Activation Function

![Graph showing the influence of activation function on system accuracy]

Fig. 10. Influence of Activation Function to The System Accuracy, the axis x denote the type of data and axis y denote the accuracy (%)

In the Fig. 10, we provide the comparison accuracy between sigmoid function and linear function. From Fig. 10 indicate the application of sigmoid function as activation function of output neuron can have the better accuracy than linear function. The difference of accuracy is 1.07% for Normalize data and 0.28% for raw data. This is happen because if it uses sigmoid function then the output neuron will be at range [0, 1]. So, it will suitable for output data of EEG Eye State dataset, that the output is 0 or 1.

C. Number of Neuron Experiment

![Graph showing the influence of number of neurons on system accuracy]

Fig. 11. Influence of Number of Neurons to The System Accuracy, the axis x denote the number of neuron and axis y denote the accuracy (%)
The result of number of neuron experiment, we present in the Fig. 11. Number of neurons were tested in this research is from 1 neuron until 50 neurons. Based on Fig. 11, it indicates that the system will have better and stable accuracy if number of neurons are more than 12 neurons. But, if the more neurons were used then the running time will become longer. Here is the result of this experiment, where the combination variables is sigmoid function, linear function, raw data, and normalize data.

**TABLE II**  
**INFLUENCE OF NUMBER OF NEURONS TO THE SYSTEM ACCURACY**

<table>
<thead>
<tr>
<th>Variables Combination</th>
<th>Number of Neuron</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sigmoid, Raw Data</td>
<td>35</td>
<td>94.62</td>
</tr>
<tr>
<td>Linear, Raw Data</td>
<td>47</td>
<td>94.44</td>
</tr>
<tr>
<td>Sigmoid, Normalize Data</td>
<td>31</td>
<td>95.71</td>
</tr>
<tr>
<td>Linear, Normalize Data</td>
<td>39</td>
<td>94.63</td>
</tr>
</tbody>
</table>

**D. Number of Hidden Layers Experiment**

**TABLE III**  
**INFLUENCE OF NUMBER OF HIDDEN LAYERS TO THE SYSTEM ACCURACY**

<table>
<thead>
<tr>
<th>Activation Function</th>
<th>Number of Hidden Layers</th>
<th>Number of Neurons</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sigmoid</td>
<td>1</td>
<td>Layer 1: 35 neurons</td>
<td>94.618</td>
</tr>
<tr>
<td>Linear</td>
<td>1</td>
<td>Layer 1: 31 neurons</td>
<td>95.707</td>
</tr>
<tr>
<td>Sigmoid</td>
<td>2</td>
<td>Layer 1: 46 neurons Layer 2: 38 neurons</td>
<td>98.01</td>
</tr>
<tr>
<td>Linear</td>
<td>2</td>
<td>Layer 1: 39 neurons Layer 2: 47 neurons</td>
<td>98.912</td>
</tr>
</tbody>
</table>

The number of hidden layers tested were one and two hidden layers with combination of 1 to 50 neurons in each hidden layer. Table III shows the experimental result of number of hidden layers. Based on the experiment shows that the highest accuracy obtains when number of hidden layers are 2 hidden layers, that is 98.912%. The difference of accuracy is 3.205% higher from result of one hidden layer. Based on this experiment, we can conclude that the more hidden layers used then the accuracy will increase, but with Levenberg Marquardt optimization the time processing will also increase significantly.

**E. Comparison Result with Standard Neural Network**

In the standard Multi-Layer Neural Network have standard learning algorithm, namely Backpropagation. Standard Backpropagation has weak convergence rate, so that it will need more time to obtain best accuracy. The standard Neural Network have been implemented to eye state detection, but its only have low accuracy. Table IV shows the comparison results between standard Neural Network and proposed scheme (Levenberg-Marquardt Neural Network).

**TABLE IV**  
**COMPARISON RESULTS BETWEEN PROPOSED SCHEME AND STANDARD NEURAL NETWORK**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Scheme</td>
<td>98.912%</td>
</tr>
<tr>
<td>Standard Algorithm (Roesler, 2013)</td>
<td>Less than 70%</td>
</tr>
<tr>
<td>Standard Algorithm (Sabanci, 2015)</td>
<td>56.45%</td>
</tr>
</tbody>
</table>

**V. Conclusion**

The proposed scheme of this research uses Encephalo Graphic (EEG) dataset to detect the eye state. The method used was Multi-Layer Neural Network with Levenberg Marquardt optimization learning algorithm. The
experiments conducted to optimize some parameters of neural network to obtain the best accuracy, that are data normalization, activation function, the number of neurons in hidden layer, and number of hidden layers. Based on the analysis towards the experiment on the EEG dataset, it can be concluded that the proposed scheme can be implemented to detect the Eye State. The best accuracy gained from combination variable sigmoid function, data normalization and number of neuron is 31 (95.71%) for one hidden layer, and 98.912% for two hidden layers with the number of neurons are 39 and 47 neurons and linear function. The results of proposed scheme with Levenberg Marquardt optimization can obtains better accuracy than standard algorithm, that is 30% higher (Rosler, 2013).

For the future work it is important to experiment with more hidden layer for the architecture of the neural network, and maybe use the learning algorithm that need less memory and less iteration, but still have good performance.

ACKNOWLEDGMENT

The authors would like to thank the Artificial Intelligence Laboratory, School of Computing, Telkom University, for support on this research.

REFERENCES


