Measuring the Accuracy of Simple Evolving Connectionist System with Varying Distance Formulas

Al-Khowarizmi¹, O S Sitompul²*, Suherman³, E B Nababan⁴
¹,² Department of Computer Science
⁴ Department of Information Technology
Faculty of Computer Science and Information Technology
Universitas Sumatera Utara, Medan – 20155, Indonesia
³ Department of Electrical Engineering
Faculty of Engineering, Universitas Sumatera Utara, Medan – 20155, Indonesia

*Corresponding Author. E-mail address: opim@usu.ac.id

Abstract. Simple Evolving Connectionist System (SECoS) is a minimal implementation of Evolving Connectionist Systems (ECoS) in artificial neural networks. The three-layer network architecture of the SECoS could be built based on the given input. In this study, the activation value for the SECoS learning process, which is commonly calculated using normalized Hamming distance, is also calculated using normalized Manhattan distance and normalized Euclidean distance in order to compare the smallest error value and best learning rate obtained. The accuracy of measurement resulted by the three distance formulas are calculated using mean absolute percentage error. In the training phase with several parameters, such as sensitivity threshold, error threshold, first learning rate, and second learning rate, it was found that normalized Euclidean distance is more accurate than both normalized Hamming distance and normalized Manhattan distance. In the case of beta fibrinogen gene -455 G/A polymorphism patients used as training data, the highest mean absolute percentage error value is obtained with normalized Manhattan distance compared to normalized Euclidean distance and normalized Hamming distance. However, the differences are very small that it can be concluded that the three distance formulas used in SECoS do not have a significant effect on the accuracy of the training results.

1. Introduction
Artificial neural networks (ANN) as one of the information processing systems are inspired from biological network of neurons and have been proven to have high performance [1]. As a modern applied computational mathematics, ANN is used to solve problem by learning process from data patterns [2]. One known method in ANN is Simple Evolving Connectionist System (SECoS) [3] which is also known as simple evolving multi-layer perceptron as a minimal implementation of Evolving Connectionist Systems (ECoS) [4]. ECoS is a computational intelligent method based on neural network, but is using a technique that differ from those of computational intelligence techniques. The ECoS operated continuously, adapting to some structure and functionality on every iterative interaction with the environment and the rest of the systems. By this manner, ECoS is able to overcome the commonly encountered ANN weaknesses in terms of longer training time and difficulty in determining the system architecture [5].
By using normalized Hamming distance in the activation function, [6] confirmed that SECoS is very sensitive to parameters such as first and second learning rate and could recognize data better and adapted faster to new data while in the learning process. Furthermore, in [4] SECoS was used to classify voice recognition by optimizing the parameter value. In the learning process normalized Hamming distance was used in the activation function to obtain smallest error. In addition to classification in medical case, [7] also used SECoS to classify gene data from cancer tissues. In [8] ECoS was implemented to develop a system which can help radiologists to identify the symptoms of breast cancer by using Breast Imaging Reporting and Database System as a standard system.

Adaptive capability of SECoS in the learning process is also studied in [9] in which the activation function is also calculated using normalized Hamming distance. Evolving Multi-layer Perceptron [10] for learning process in big data using normalized Hamming distance in the activation function obtained smallest error rate. Furthermore, using a different method, [11] compared Manhattan distance and Euclidean distance as two most popular distance formulas and came to a conclusion that Euclidean distance had more iteration compared to Manhattan distance. To optimize discrete wavelet frame (DWF) configuration, [12] used texture feature extraction method on images which used Manhattan distance, Euclidean distance, normalized Manhattan distance and normalized Euclidean distance. They concluded that normalized Euclidean distance gave an optimal value.

Various applications of SECoS had been conducted in the field of medical, environment and health studies [7]. In this research, a calculation of the activation value in SECoS is performed not only using normalized Hamming distance formula but also using normalized Manhattan distance and normalized Euclidean distance in order to compare the smallest error value and the best learning rate obtained on the case of beta fibrinogen gene -455 G/A polymorphism patients. The accuracy of the SECoS training results from the three distance formula is then calculated using mean absolute average error (MAPE) as one of the accuracy measure commonly used to calculate percentage error in statistical forecasting [13].

2. Methodology

2.1. Dataset Used

Data used in this research is data from beta fibrinogen gene -455 G/A polymorphism patients obtained from the examination of neurological and fibrinogen concentration rate [14]. The data belong to 136 patients which have 9 characteristics such as age, Hb, Ht, Leucosite, KGD, cholesterol, HDL, LDL and fibrinogen concentration rate.

2.2. General Architecture

The general architecture of this research is depicted in Figure 1. Data from the database is firstly normalized in the range of 0-1. These data are then given to the SECoS training network, whereby each of the distance formula is used in the calculation. Lastly, The accuracies of the output of the SECoS network are calculated using the mean absolute percentage error (MAPE).

Figure 1. The general architecture
2.3. Simple Evolving Connectionist Systems

SECoS consists of three neuron layers [6], in which the first layer is an input layer with linear or other activation functions, the second layer is an evolving hidden layer, and the third layer is the output layer. The linear activation function $A_n$ on the evolving node $n$ is defined by [4] as in (1).

$$A_n = 1 - D_n$$  

(1)

where $A_n$ is the activation value on node $n$ and $D_n$ is normalized distance value between input vector and incoming weight vector on corresponding node. In general, distance value $D_n$ is calculated using the following normalized Hamming distance as suggested in [4]:

$$D_n = \frac{\sum_{i=1}^{K} |I_i - W_i|}{\sum_{i=1}^{K} |I_i + W_i|}$$

(2)

where $K$ is the number of input nodes on SECoS, $I$ is the input vector, and $W$ is the input weight matrix on evolving layer.

In this paper, an analysis on the activation function is performed with distance values are calculated not only using normalized Hamming distance but also using normalized Manhattan distance and normalized Euclidean distance. Formula to calculate the distance values as adapted from [12] are shown in (3) and (4), respectively.

$$D_n = \frac{1}{k} \left( \sum_{i=1}^{K} |I_i - W_i| \right)^2$$

(3)

$$D_n = \frac{1}{k} \left( \sum_{i=1}^{K} |I_i - W_i| \right)^2$$

(4)

A learning process is then conducted using the SECoS method on patients data of beta fibrinogen gene -455 G/A polymorphism. The SECoS algorithm is adapted from [15] as in the following:

Step 1. Propagate input vector $I$ into the network.

Step 2. If the maximum activation ($A_{\text{max}}$) from one node is less than the sensitivity threshold ($S_{\text{thr}}$) coefficient, then add a new node. Otherwise,

a. Calculate the error value between the learning output of output vector $O_c$ and the actual value of output vector $O_d$.

b. If the error value is greater than the error threshold ($E_{\text{thr}}$) coefficient or output node desired is not active, then:

   i. Add a new node, otherwise

   ii. Change the connection weight on the winning hidden node.

Step 3. Repeat Step 1 and Step 2 for each input vector.

Whenever a node is added, the input weight is initialized correspond to input vector $I$ and the output weight is initialized corresponds to output vector $O_d$. Propagation process from hidden layer to output layer can be performed in two ways. First is using One-of-N propagation method in which propagation is performed by the hidden node with highest activation value. Second is using Many-of-N propagation method where propagation is performed by hidden node with activation value above the activation threshold.

2.4. Mean Absolute Percentage Error (MAPE)

The amount of result errors are calculated using MAPE as shown in (5). Among various accuracy criteria, MAPE is a measuring tool which incorporating the best characteristics [13]. Apart from some
its disadvantages, MAPE is a popular and widely used accuracy measure used in businesses and organizations [16].

\[
MAPE = \frac{1}{n} \left( \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \right) \times 100
\]  

(5)

where \(A_t\) is the actual data, \(F_t\) is calculated data, and \(n\) is the amount of data.

3. Result and Discussion

In this section, result of comparing normalized Hamming distance, normalized Manhattan distance and normalized Euclidean distance on the SECoS activation function will be examine to obtain the smallest error value and the best learning rate in the training stage.

3.1. Test results with three distance formulas

In performing a training stage using SECoS method, the first thing to do is determining the parameter values, namely learning rate 1, learning rate 2, sensitivity threshold, error threshold and the number of training data. The parameter values are then normalized in the range of 0-1, continued by propagation from input layer to hidden layer by calculating activation values using normalized Hamming distance and finally performing propagation from hidden layer to output layer.

The activation functions of normalized Hamming distance, normalized Manhattan distance and normalized Euclidean distance in the SECoS training are tested using beta fibrinogen gene -455 G/A polymorphism patient data. Testing for those distance formulas will be compared using various combination of parameter values. The first input parameter values are calculated with sensitivity threshold = 0.6, error threshold = 0.1, learning rate 1 = 0.3, and learning rate 2 = 0.3, resulted in a MAPE value of 0.56618 and 24 hidden nodes, 0.57179 MAPE value and 24 hidden nodes, and 0.60317 MAPE value and 28 hidden nodes, for normalized Hamming distance, normalized Manhattan distance, and normalized Euclidean distance, respectively (see the first row of Table 1). As the parameter values used on the training stage are affecting the resulted error values, therefore some combination of the parameter values are performed in order to obtain the smallest error. The complete combination of parameter values can be seen in Table 1.

| Test No | Sensitivity Threshold | Error Threshold | Learning Rate 1 | Learning Rate 2 | MAPE  | Hidden Node | MAPE  | Hidden Node | MAPE  | Hidden Node |
|---------|----------------------|-----------------|-----------------|-----------------|-------|-------------|-------|-------------|-------|-------------|
| 1       | 0.6                  | 0.1             | 0.3             | 0.3             | 0.56618 | 24          | 0.57179 | 24          | 0.60317 | 28          |
| 2       | 0.6                  | 0.1             | 0.3             | 0.6             | 0.57644 | 27          | 0.59153 | 33          | 0.66081 | 33          |
| 3       | 0.6                  | 0.1             | 0.3             | 0.9             | 0.58201 | 34          | 0.60809 | 34          | 0.63439 | 37          |
| ...     | ...                  | ...             | ...             | ...             | ...    | ...         | ...    | ...         | ...    | ...         |
| 19      | 0.3                  | 0.1             | 0.3             | 0.3             | 0.52377 | 22          | -       | -           | -       | -           |
| ...     | ...                  | ...             | ...             | ...             | ...    | ...         | ...    | ...         | ...    | ...         |
| 22      | 0.3                  | 0.1             | 0.6             | 0.3             | -      | -           | 0.52657 | 25          | 0.51691 | 24          |
| ...     | ...                  | ...             | ...             | ...             | ...    | ...         | ...    | ...         | ...    | ...         |
| 34      | 0.3                  | 0.05            | 0.9             | 0.3             | 0.55399 | 27          | 0.55280 | 29          | 0.54024 | 28          |
| 35      | 0.3                  | 0.05            | 0.9             | 0.6             | 0.56651 | 33          | 0.58345 | 36          | 0.59051 | 35          |
| 36      | 0.3                  | 0.05            | 0.9             | 0.9             | 0.55603 | 36          | 0.57230 | 37          | 0.54533 | 37          |

Table 1. Combination of parameter values using three distance formulas
It is shown in Table 1 that the smallest error obtained in the SECoS training with activation function using normalized Hamming distance obtained at test number 19 where sensitivity threshold = 0.3, error threshold = 0.1, learning rate 1 = 0.3, and learning rate 2 = 0.3, with a MAPE value of 0.52377 and 22 hidden nodes. Training stage with the activation function using normalized Manhattan distance resulting in the smallest error at test number 22 where sensitivity threshold = 0.3, error threshold = 0.1, learning rate 1 = 0.6, and learning rate 2 = 0.3, with a MAPE value of 0.52657 and 25 hidden nodes. Using normalized Euclidean distance the smallest error obtained at test number 22 where sensitivity threshold = 0.3, error threshold = 0.1, learning rate 1 = 0.6, and learning rate 2 = 0.3, with a MAPE value of 0.51691 and 24 hidden nodes.

3.2. Discussion

On the test with normalized Hamming distance, it can be observed that the smaller sensitivity threshold value used, the smaller the error value obtained. In term of error threshold, the larger the value is given, the smaller the error value obtained. The same pattern is also hold for the learning rate 1 and learning rate 2, that is the smaller the values given, the smaller the error value. On the sensitivity threshold and error threshold the same patterns are also obtained from the test on normalized Manhattan distance and normalized Euclidean distance. However, in normalized Manhattan distance the smaller the value of learning rate 1 and learning rate 2 used, the smaller the error value obtained, whereas with normalized Euclidean distance the smaller learning rate 1 and learning rate 2 used, the smaller the error value obtained.

Result of parameter combination and test with normalized Hamming distance as shown in Table 1, give the smallest error value of 0.52377 obtained with a combination of parameter learning rate 1 = 0.3, learning rate 2 = 0.3, sensitivity threshold = 0.3 and error threshold = 0.1. As for the largest error value obtained is 0.63086 with combination of parameter learning rate 1 = 0.9, learning rate 2 = 0.6, sensitivity threshold = 0.6 and error threshold = 0.05. Meanwhile, result of combination and test with normalized Manhattan distance shows that the smallest error obtain is 0.52657 using parameter combination of learning rate 1 = 0.6, learning rate 2 = 0.3, sensitivity threshold = 0.3 and error threshold = 0.1. As for the largest error value obtain is 0.68776 with combination parameters learning rate 1 = 0.9, learning rate 2 = 0.9, sensitivity threshold = 0.6 and error threshold = 0.1. Finally, parameter combination and test using normalized Euclidean distance shows the smallest error value of 0.51691 with combination of parameters learning rate 1 = 0.6, learning rate 2 = 0.3, sensitivity threshold = 0.3 and error threshold = 0.1. The largest error obtained is 0.67534 with combination parameters learning rate 1 = 0.3, learning rate 2 = 0.6, sensitivity threshold = 0.6 and error threshold = 0.05. Summary of the test result on the beta fibrinogen gene -455 G/A polymorphism patient data is given in Table 2.

| No | Distance Formula       | MAPE   | Parameter Value                  |
|----|------------------------|--------|----------------------------------|
|    |                        |        | Sensitivity Threshold | Error Threshold | Learning Rate 1 | Learning Rate 2 |
| 1  | Normalized Hamming     | 0.52377| 0.3     | 0.1                        | 0.3          | 0.3          |
|    | distance               |        |         |                            |              |              |
| 2  | Normalized Manhattan   | 0.52657| 0.3     | 0.1                        | 0.6          | 0.3          |
|    | distance               |        |         |                            |              |              |
| 3  | Normalized Euclidean   | 0.51691| 0.3     | 0.1                        | 0.6          | 0.3          |
|    | distance               |        |         |                            |              |              |

Table 2. Comparison of Distance Formulas
4. Conclusion
The smallest MAPE value in the test using beta fibrinogen gene -455 G/A polymorphism patient data is obtained using normalized Euclidean distance. However, the differences is very small that it can be concluded that the three distance formulas used in SECoS do not have a significant effect on the accuracy of the training results. The smallest error on the SECoS training is also affected by sensitivity threshold and error threshold. In term of hidden nodes generated, the number of nodes are always different since SECoS algorithm perform the training continuously based on the input data given.

Acknowledgement
The author would like to thanks Lembaga Penelitian Universitas Sumatera Utara (LP USU) for partly supporting the dissemination of this research.

References
[1] Aram Z, Jafari S, Ma J, Sprott JC, Zendehrouh S and Pham V-T 2017 Using chaotic artificial neural network to model memory in the brain Comm Nonlinear Sci Numer Simulat 44 449–459
[2] Khaze SR, Masdari M and Hojjatkhah S 2013 Application of artificial neural networks in estimating participation in election. Int. J. of Technology Information, Modelling and Computing 1 23–32
[3] Watts MJ 2006 Nominal-scale evolving connectionist systems Proc. of the 2006 IEEE Int. Joint Conf. on Neural Network 2055–59
[4] Watts MJ and Kasabov NK 2002 Evolutionary optimisation of evolving connectionist systems. Proc. of the Congress on Evolutionary 606–9
[5] Kasabov NK 2007 Evolving Connectionist System The Knowledge Engineering Approach. 2nd Ed. Springer: London
[6] Watts MJ and Kasabov NK 2000 Simple evolving connectionist systems and experiment on isolated phoneme recognition Proc. of the 2000 IEEE Symposium on Combinations of Evolutionary Computation and Neural Network 232–9
[7] Futschik ME, Reeve A and Kasabov N K 2003 Evolving connectionist systems for knowledge discovery from gene expression data of cancer tissue. Artificial Intelligence in Medicine 28 165–189
[8] Nababan EB, Iqbal M and Rahmat RF 2016 Breast cancer identification on digital mammogram using evolving connectionist systems 2016 IEEE International Conference on Informatics and Computing (ICIC) 132–6
[9] Gholakhlou A, Watts MJ and Kasabov NK 2003 Adaptive speech recognition with evolving connectionist systems Information Science 156 71–83
[10] Rahmat RF, Rizki A, Alharthi AF and Budiarto R 2016 Big data forecasting using evolving multi-layer perceptron 2016 4th Saudi Int.Conf. on Information Technology 1–5
[11] Sinwar D and Kaushik R 2014 Study of Euclidean and Manhattan distance metrics using simple k-means clustering Int. J. for Res.in Applied Science and Engineering Technology 2 270–4
[12] Fauzi MFA 2009 Optimal Discrete Wavelet Frames Features for texture-based image retrieval applications Proc. of the First International Visual Informatics Conference 66–77
[13] Makridakis S 1993 Accuracy measures: theoretical and practical concerns Int J.of Forecasting 9 527–9
[14] Ritarwan K 2014 Effect of beta fibrinogen gene -455 G/A polymorphism on aspirin treatment and its effect to the barthel index score and modified ranking scale on ischemic stroke based on the age groups. PhD Dissertation. Universitas Sumatera Utara
[15] Watts MJ 2016 Sleep learning and max-min aggregation of evolving connectionist systems International joint conference on Neural Network (IJCNN) 4325–30
[16] Tofallis C 2014 A better measure of relative prediction accuracy for model selection and model estimation Journal of Operational Research Society 1–11