Neural Networks Approach in Medical Analysis

Shuzlina Abdul Rahman
School of Information Technology
Universiti Utara Malaysia
Sintok, 06010, Jitra,
Kedah Darul Aman, Malaysia
shuzlina@uum.edu.my
Tel: 604 7003763
Fax: 604 7005753

Fadzilah Siraj
School of Information Technology
Universiti Utara Malaysia
Sintok, 06010, Jitra,
Kedah Darul Aman, Malaysia
fad173@uum.edu.my
Tel: 604 7003715
Fax: 604 7005753

Extended Abstract

This paper presents an empirical evaluation on medical, particularly classification of anaemia patients using neural networks approach. The number of hidden units, learning rate and momentum are varied, so that more appropriate classification model is obtained. The information regarding anaemia cases were collected from haematology form. In addition, seventeen attributes were indentified and used in the training model.

A total of seven hundred raw data of anaemia patients has been collected and preprocessed that includes data cleansing, data selection and data preprocessing. Data is partitioned into three data sets namely training set (80%), testing set (10%) and validation (10%) set. The training data is used to train the model while the validation data is used to monitor neural network performance during training. The test data is used to measure the performance of a trained model. The motivation here is to validate the model on a data set that is different from the one used for parameter estimation. In this way, we may use the training set to assess the performance of various candidate models, and thereby choose the "highest" one. In this exploratory study, the onset of overfitting is identified by cross-validation method that is referred to as the early stopping method of training (Haykin, 1999).

Several phases involved in investigating the network generalization when different parameter is applied. In the first phase, we investigate the effect on network generalization when the input size is fixed but the hidden units size are varied. The criterion used in selecting the hidden unit is the highest average test and lowest average training correctness. The networks have the following architecture: 17: m_h: 8. In this study, the experiment was simulated with 10 different hidden units, starting from 8 hidden units up to 17 hidden units. The best two hidden units are then retrained with different weight seeds that value from 1 to 10. This is to ensure that the selected value give a better generalization than the other one.

In the second phase, we investigate the network generalization of the multilayer perceptron with the hidden unit size fixed, m_{opt} =15 but the learning rate, α (also known as learning coefficient) is varied from 0.1 to 1.0. The same neural networks architecture is used (17: m_h:8) but the value of hidden units (m_h) is fixed to 15. The momentum constant μ is arbitrarily sets to 0.1. The same criterion is used as in the above phase, in which the learning rate that has the average highest test correctness is selected. No large differences is observed from α = 0.1 to 0.5.

In the third phase, we investigate the training behaviour of the networks with the hidden units size (m_{opt} =15) and learning rate fixed but the momentum μ is increased. Momentum with highest test correctness is selected. The networks still used the same architecture as in the previous section 17:15:8. On the average, the results show that the generalization decreases as the value of μ increases.

In the next phase, we examine the network performance when applied with different mode of weight update and distributions. The same networks architecture of 17:15:8 are used and the optimal number of hidden unit, values of learning rate and momentum that gained from the previous phase are applied to the network. Uniform and Gaussian distribution functions in Multilayer Perceptron training are observed and best correctness is achieved through Uniform function. We examine the network generalization with epoch and pattern weight updates. The results show better performance with epoch updates.

In the following phase, we examine the networks performance on different type of activation functions namely, Sigmoid, Tanh and Linear activation function. Sigmoid activation function is preferred in the beginning of the training study as it meets the nature of the problem that has data sets of real values. Using the same network architectures of 17:15:8, the behaviour of the three activation functions was observed. The performance of networks in training and testing data sets of sigmoid activation functions is better than hyperbolic Tangent activation functions. Thus supporting the statement made by Haykin (Haykin, 1999) and other researchers why sigmoid is the most preferred form of activation function. Linear activation function does not give any performance value. The results indicate that the non-linear neural networks are suitable for the research domain.

Finally we examine the performance of Multilayer perceptron (MLP) and other neural network models namely, Radial Basis Function (RBF) as well as statistical model such as Regression (REG) model. With architecture of 17:15:8, the network is set with the optimum parameters gained from all phases. The result of MLP is superior to the other models for both data sets. When compared with RBF and REG models, a difference of 7.81% and 10.62% in generalization was observed respectively. The Regression model performance is the lowest compared to neural networks model (Multilayer Perceptron and Radial Basis Function). Similar results are obtained when multilayer perceptron model used linear activation functions. These two results indicate that non-linear model such as multilayer perceptron and radial basis function are able to generalize the problem domain and multilayer perceptron has shown to be the most suitable model in the context of this study.

The results of the study show that the multilayer perceptron (MLP) model used for predicting the classification for anaemia with backpropagation learning rule can be achieved by applying the optimal value of number of hidden units, learning rate, and momentum to the selected patient. In addition, the model achieved higher generalization compared to Radial Basis Function (RBF) and Regression model when sigmoid activation function was employed. The highest performance was obtained when the number of hidden units is 15, learning rate is 0.7 and momentum is 0.1. The testing and generalization correctness is 71.56 and 72.78 respectively. This result has demonstrated the ability of multilayer perceptron for predicting classes of anemia and can be used by haematologist and other medical staff.

Keywords: Medical, Anaemia, Backpropagation, Radial Basis Function, Regression