British Journal of Mathematics & Computer Science
18(4): 1-14, 2016, Article no.BJMCS.28409
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

Cascade Backward Propagation Neural Network and Multiple Regression in the Case of Heteroscedasticity

Mamman Mamuda[1*] and Saratha Sathasivam[1]

[1] *School of Mathematical Sciences, University Science Malaysia, 11800 Pulau Pinang, Malaysia.*

Authors' contributions

This work was carried out in collaboration between both authors. Author MM designed the study, performed the statistical analysis, wrote the protocol, and wrote the first draft of the manuscript and managed literature searches. Authors MM and SS managed the analyses of the study and literature searches. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJMCS/2016/28409
Editor(s):
(1) Kai-Long Hsiao, Taiwan Shoufu University, Taiwan.
Reviewers:
(1) Radosaw Jedynak, Kazimierz Pulaski University of Technology and Humanities, Poland.
(2) Loc Nguyen, Sunflower Soft Company, Vietnam.
(3) Rajesh Chandrakant Sanghvi, Gujarat Technological University, Gujarat, India.
(4) Utku Kose, Usak University, Turkey.
Complete Peer review History: http://www.sciencedomain.org/review-history/15994

Received: 18th July 2016
Accepted: 16th August 2016
Published: 31st August 2016

Original Research Article

Abstract

Aims/ Objectives: To develop a new model called cascade backward propagation neural network performance over a filtered data by clustering algorithm based on robust measure ($CFBNFDCARM$). The performance of the clustering based neural network approach will be compare with the performances of regression analysis when the data deviate from the assumption of homoscedastic regression.

Methodology: The new developed model was tested using the Airfoil, Aboline and Airline passenger data sets obtained from the UCI machine learning repository in order to compare the performances of regression analysis and a clustering based neural network approach when the data deviate from the assumption of homoscedastic regression. An algorithm based on robust estimates of location and dispersion matrix that helps in preserving the error assumption of the linear regression was introduced in the clustering technique.

*Corresponding author: E-mail: maanty123@gmail.com;

Results: The comparison indicated that the results emerging from our developed model gives a better performance when compared with the weighted least square regression as well as the standalone cascade backward propagation neural network for all the data sets considered.
Conclusion: Analysis of the result showed that, the mean square error (MSE) and the root mean squared error ($RMSE$) in all the cases considered in this study decreases in a definite manner. From the obtained result, it can be seen that, our proposed model ($CBPNFDCARM$) performed better and can be a better alternative in dealing with heteroscedasticity in data set than both the weighted least square (WLS) and the standalone cascade backward propagation neural network ($CBPN$).

Keywords: Cascade backward propagation neural network; heteroscedasticity; regression analysis; robust estimate; clustering algorithm.

2010 Mathematics Subject Classification: 53C25, 83C05, 57N16.

1 Introduction

In prediction analysis, linear regression is found to be a classical and commonly used prediction tools. However, it demands data that satisfy certain criterion such as linearity, additivity and homoscedasticity. In complex data sets, the said criterion may not be satisfied. Hence, the development of modern regression and other techniques with the capability of handle the prevalence challenges in complex data sets that includes non-linearity and non-additivity. Research on how heteroscedasticity in data influences the predictive ability of regression and other models is very sparse. Various and classical methods of detecting outliers were presented and subjected to limitations. [1], [2] proposed a modified identification of outliers in multivariate data set using a robust estimate in detecting outliers of a data set. His approach yielded to controlling the size of the test which led to an improved power and effective in dealing with the problem of masking and swamping. Nag et al. [3] proposed an artificial intelligence technique with the use of self-organizing map (SOM) for detecting multiple outliers in multidimensional data sets. Ali and Ong [4], used a minimum mohalanobis distance (MMD) to construct a cluster phase to a bounded influence regression phase in their propose robust regression method of identifying outliers. Findings from their method showed that, the resulting proposed method has a greater advantage over other robust regression techniques.

Carroll and Ruppert [5] assert heteroscedasticity as a very common observed problem in regression analysis which in turn results in increasing the variance of parameter estimation and thus affects coefficient of determination (R^2), estimated (σ^2) and other inferential procedures accordingly. In data sets, the frequently existence of non- constant variability is a general phenomenon that is encountered in nearly all fields with many different forms of heteroscedasticity Daye et al. [6], Brem and Kruglyak [7], Lim et al. [8]. Various form of approaches that deal with the problem of non-constant error variance in regression analysis are broadly discussed. A research by Montgomery et al. [9] and Gujarati et al. [10] suggested an approach to deal with non constant error variance of a data in regression analysis by applying a weighted least square (WLS) based upon the assumption that, the non-constant error variance is known up to a constant of proportionality. Ong and Alih [11] assert that, diagnosis of heteroscedasticity in data set can be done using both graphical and inferential procedures which are very much available in literature. However, graphical procedures were always exposed to errors and the test to detect heteroscedasticity have many limitations. Pwasong and Saratha [12] developed a method known as the hybrid quadratic regression and cascade forward backpropagation neural network ($QRM - CFBN$) using the bayesian averaging model technique in forecasting the performance of the joint integration. Log difference series was applied to the original time series data. The result obtained from their hybrid model showed that the ($QRM - CFBN$) forecasting performance generally outperform the standalone model. The assumption of the presence of outliers in data which often violet the assumption of both normality and homoscedasticity was ignored from their study.

In this study, we proposed a method to handle the deviation of data set from the assumption of homoscedasticity leading to heteroscedasticity. A cascade backward propagation neural network over a filtered data by clustering algorithm based on robust measure ($CBPNFDCARM$) was proposed with the intention of addressing the cause of heteroscedasticity in data sets. Presence of contamination or outliers in data set is very common in practice, and makes a homoscedastic model heteroscedastic. This study tends to eliminate the outliers from a data set using a cluster algorithm and then fits in a cascade backward propagation neural network whose structure is determined using the normality assumption of linear regression. The cluster algorithm considered in this study used the minimum mohalanobis distance (MMD) and robust estimate to define the radius of the cluster. This motivated us to conduct the present study toward preserving the assumptions of linear regression. To gain insight on how the comparative performance of our technique and other techniques under heteroscedasticity translate to real life situations, data set were also been analyzed and results were discussed.

The remaining part of this research is organized as follows: Weighted least square regression as well as cascade backward neural network were summarized in section 2. Methodology and our new developed technique described in section 3. which is an integration of a clustering algorithm and neural network. Results analysis as well as findings on the outcome of the results that arises from the pragmatic findings of the study were presented in section 4. Conclusion, remarks as well as future work are presented in section 5.

2 Weighted Least Square

Consider the multiple linear regression model given by:

$$Y_i = \beta_o + \beta_1 X_{1i} + ... + \beta_j X_{ji} + ... + \beta_p X_{pi} + \varepsilon_i \forall i = 1, 2, ..., n; j = 0, 1, 2, ..., p + 1 \tag{2.1}$$

where Y_i is the dependent variable, $X_{j,i}$ are independent variables, β_j are unknown parameters and ε_i is the error term Consider the following assumptions of a Multiple linear regression (MLR):

(i) $E(\varepsilon_i) = 0$; for all i
(ii) $V(\varepsilon_i) = \sigma^2$; for all i
(iii) $E(\varepsilon_i \varepsilon_j) = 0$; whenever $i \neq j$

Satisfying all the above assumption means OLS as the best linear minimum variance estimate among the class of unbiased estimators. If assumption (ii) is violated, then there exist a case of heteroscedasticity. In this case, OLS estimates are no longer the best linear unbiased estimator, though are still unbiased under heteroscedasticity. Weighted least square (WLS) is an estimation procedure that is usually followed whenever errors in a regression model are uncorrelated but do not have equal variances. The estimates of the parameters of a weighted least square are obtained by minimizing the equation below.

$$\sum_{i=1}^{n} w_i (Y_i - \beta_o - \beta_1 X_{1i} - \beta_2 X_{2i} - ... - \beta_p X_{pi})^2 \tag{2.2}$$

The weight W_i is inversely related to the variance σ^2 which reflects the amount of information contained in Y_i. Hence an observation Y_i that has a large variance receives less weight than other observation that has a smaller variance. The weighted least squares estimates of the regression coefficients can easily be expressed in terms of matrix as follows.

$$Y_{n\times 1}=\begin{bmatrix} Y_1\\ Y_2\\ \vdots\\ Y_n\end{bmatrix} \quad ; \quad X_{n\times(p+1)}=\begin{bmatrix} 1 & X_{11} & \cdots & X_{1p}\\ 1 & X_{21} & \cdots & X_{2p}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & x_{n1} & \cdots & X_{np}\end{bmatrix} \quad ; \quad W=\begin{bmatrix} w_1 & 0 & \cdots & 0\\ 0 & w_2 & \vdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & w_n\end{bmatrix}$$

With a given diagonal matrix W containing weights w_i.

The weighted least square regression coefficients in the form of the normal equation can now be represented as:

$$\left.\begin{aligned}(X^tWX)b_w = X^tWY\\ b_w=(X^tWX)^{-1}X^tWX\end{aligned}\right\} \tag{2.3}$$

Weighted least square estimated regression coefficient with variance co-variance matrix is represented as:

$$\sigma^2(b_w)=\sigma^2(X^tWX)^{-1} \tag{2.4}$$

Equation (2.5) is a matrix that is usually not known as the proportionality constant σ^2 is rarely known. In this case however, consistent estimate of σ^2 can be use which is estimated as:

$$\left.\begin{aligned}S^2\{b_w\}=S_w^2(X^tWX)^{-1}\\ S_w^2=\sum_{i=1}^{n} w_i\frac{(Y_i-\hat{Y}_i)^2}{(n-p-1)}\end{aligned}\right\} \tag{2.5}$$

where S_w^2 is based on weighted squared residuals. On a transformed model, weighted least square can also be seen as ordinary least square given by:

$$Y_w = X_w\beta+\varepsilon_w \tag{2.6}$$

2.1 Cascade backward propagation neural network ($CBPN$)

Neural network which is seen as an information processing system that is designed to model the capability of biological neurons of human like brain and a well known technique with the ability to estimate functional relationship were used for prediction/forecasting analysis. Significant contributions of neural networks in the field of prediction problems that have put the field on a strong theoretical and conceptual foundation were carried out. The comparison between the performance of a feed forward neural network and linear regression was presented by Paliwal and Kumar [13] with an emphasis that the data tend to deviate from assumptions of homoscedasticity of regression analysis. In estimating the parameters of their model, Weighted least square method (WLS) was applied. Findings from their study further revealed that, based on the fact of the capability of neural network in estimating functional relationship, feed forward neural network outperformed the weighted least square method. The drawback in their prediction model is that, the author ignored the presence of outliers in the data set which may disagree with the assumption of normality or even both normality and homoscedasticity. In our study, the new method ($CBPNFDCARM$) is used to address that assumptions.

According to Thatoi et al. [14] a cascade backward propagation neural network is a network that has same characteristic as the feed forward back propagation neural network. The only differences that exist between the two is that, the input values of the cascade backward propagation neural network ($CBPN$) are computed after every hidden layer and are back propagated to the input layer and adjust the weight successively. The input values of the network are directly connected to the final output and the existence of the association between the values obtained from the hidden layers and the values obtained from the input layers as well as the consequently adjustment of the weights. Where as in the feed forward back propagation neural network ($FFBNN$), the network can effectively learn any possibly input-output relationship with more layers that might learn multifarious relationship. Singh and Srivastava [15] proposed a cascade neural network for the contingency and ranking of line flow that was performed to choose the contingencies that cause the worst overloading problems. Result from their findings showed that, the trained cascade forward neural network was able to perform the task of screening and rank all the critical contingencies correctly and in accordance to their severity. They further compared performance of the cascade neural network with four layered feed forward artificial neural network and found that, the trained cascade neural network gives an accurate and fast screening as well as ranking for unknown patterns. Hedayet et al. [16], on estimation of research reactor core parameters, affirmed that the pattern of the core reload program as a very important factor for optimizing the uses of research reactor. They also reported that, cascade feed forward neural networks improves effectively the pattern optimization process of core reload program. Fig. 1. depict the structure of the cascade forward back propagation neural network.

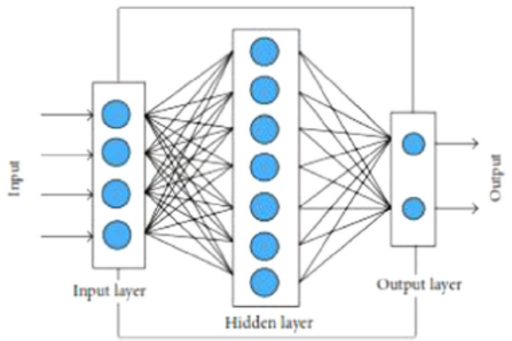


Fig. 1. Structure of the cascade forward back propagation neural network architecture

Because of the unique features of the cascade backward propagation neural network, i.e the input values of the cascade backward propagation neural network are being computed after every hidden layer and back propagated to the input layer and successive adjustment of the weights. The researchers tend to introduce cascade backward propagation neural network in determining the performance of the network in case of heteroscedastic data sets.

3 Methodology

The homogeneity of residual variance in ordinary least squares (OLS) allows an easy computation and form a close solution that enjoys the minimum variance property. It is often applied in the field of engineering and applied sciences. Research on the cause at which the assumption of homoscedastic error variance breaks down to set in heteroscedasticity were elucidated, among which are the works of Carrol and Ruppert [5] and Rana et al. [17]. In classical regression theory, weighted least square (WLS) is one of the methods used in dealing with heteroscedasticity. The existence of outliers in a data set are presumed to make a model to deviate from the assumption of homoscedasticity. According to Paliwal and kumar [13]; the performance measure of neural networks with regression analysis when compared using a simulated data set showed that the errors obtained in training neural network are generally smaller than errors obtained in weighted least square regression analysis with more differences from the smaller sample size and becoming less for large sample size. Atkinson and Riani [18] assert that, the procedure of most outlier detection tends to divide the data into two parts, a part that contain the removed outliers (clean part) and the part that contain the outliers (outlier part). The clean part are then used for parameter estimations. Our method followed the approach of both Paliwal [13] and Atkinson [18]. But our method lay more emphasis by firstly removing the outliers from data set using a clustering algorithm based on robust measures acronym as ($CBPNFDCARM$). Once the outliers are removed, cascade backward neural network is fitted to the remaining data set using back propagation learning algorithm for training processes. Classical methods of outliers detection are powerful when data contain only one outlier. However, the power of these methods does not work properly when more than one outlier observation are present in the data. Robust estimate of mean and covariance matrix were used in defining the radius of the clustering algorithm since these estimates are free from the problem of masking and swamping.

3.1 The new developed $CBPNFDCARM$ model

The equation of the new developed $CBPNFDCARM$ model is obtained as follows:

$$Y_i = \hat{Y}_i + g(x) \tag{3.1}$$

Where $\hat{Y}_i$ is the Minimum Mahalanobis Distance (MMD) and g(x) is the feed forward neural network. Consider a data set $m = m_{y,i} := (x_i, y_i); i = 1, 2, ..., n \subset \Re^k$ then

$$\hat{Y}_i = A_s(m, p_j, C^{-1}) = \sqrt{(m - p_j)^t C^{-1}(m - p_j)} \tag{3.2}$$

where,

$A_s(m, p_j, C^{-1})$ denotes the robust Mahalanobis distance based on the classical mean and convariance matrix exclusively.

If $p_j \in \Re^{1\times k}$ and $C^{-1} \in \Re^{k\times k}$; then MMD estimate can be define as:

$$p_j(H, m) = \frac{1}{h}\sum_{j\in H} m \tag{3.3}$$

and,

$$C^{-1}(H, m, p_j) = \frac{1}{h}\sum_{j\in H}(m - p_j)(m - p_j)^t \tag{3.4}$$

The squared Mahalanobis distance with respect to p_j and C^{-1} is also define as:

$$A_s^2(m, pj, C^{-1}) = \{(m - p_j)^t C^{-1}(m - p_j)\} \tag{3.5}$$

$A^2_{1:n}(m, p_j, C^{-1}) \leq ... \leq A^2_{n:n}(m, p_j, C^{-1})$ as the ordered sequence of the distance in equation (3.5). Hence

The estimator of MMD can be obtained as follows:

$$Argmin \sum_{j=1}^{h} A^2_{j:n}(m, p_j, C^{-1}) \tag{3.6}$$

where p_j and C^{-1} are the location and dispersion estimators based on the subsample. Hence;

$$\{MMD|(m, \tilde{p_j}, \tilde{C^{-1}}) \in Argmin \sum_{j=1}^{h} A^2_{j:n}(m, p_j, C^{-1})\} = \{MMD(\hat{H})|\hat{H} \in Argmin A^2_{j:n}(m, p_j, C^{-1});\tag{3.7}$$

This shows that, any (p_j, C^{-1}) minimizing the sum of h smallest squared mahalanobis distances based on the subsample is also a solution to equation (3.4) and (3.5). The objective function of the MMD estimator can therefore be define as:

$$MMD(m) = {}^{Argmin}_{HC_{-1}} A^2_{j:n}(m, p_j, C^{-1}) \tag{3.8}$$

The output neurons of the cascade forward network is given by the sigmoid function written as:

$$y = f(x, w) = \frac{1}{1 + exp(-w_o - \sum_i^m w_i x_i)} \tag{3.9}$$

From equation (3.9), the cascade forward equation can therefore be deduce as follows:

$$g(x) = \sum_i^p \psi_i [1 + exp(-w_o - \sum_i^m w_i x_i)]^{-1} + \varepsilon_i \tag{3.10}$$

where ψ_i is the connection weight that connect the output layer neuron to the hidden layer neuron, p stands for the number of neurons present in the hidden layer,w_i is a $q \times 1$ weight vector, w_o is the bias term and $x_i (i = 1, ..., m)$ is the $m \times 1$ input vector respectively.

The equation of the new developed model i.e. $CFBNFCARM$ can now be express as:

$$Y_i = A_s(m, p_j, C^{-1}) = \sqrt{(m - p_j)^t C^{-1} (m - p_j)} + \sum_i^p \psi_i [1 + exp(-w_o - \sum_i^m w_i x_i)]^{-1} + \varepsilon_i \tag{3.11}$$

Our method which is referred to as cascade backward propagation neural network performance over a filtered data by clustering algorithm based on robust measure ($CBPNFDCARM$) has the following distinct features that differentiate it from any current neural network performance measures obtainable in the literature.

1. The outliers were firstly removed from the data using the clustering algorithm, since outliers in data leads to a model to deviate from the assumption of homoscedasticity.
2. The clustering algorithm tends to divide the data set into two part, a part that contains the removed outliers (clean part) and the part that contains the outliers (outlier part).
3.The clean part of the data i.e. the removed outliers part were then fitted into the neural network for training using the back propagation learning algorithm in order to determine the measure of the performance of the network.

Training or learning algorithm or rule is a procedure to modify the weights and biases of a network. The training algorithm adjusts the weights and biases to move the network output closer to the target. The proposed method is illustrated via the following algorithm:

Clustering Algorithm

Step 1. Input $(n \times p)$ dimensional data set, that is A.

Step 2. Find the initial shape of the cluster, that is, $C_i (i = 1, 2, ...)$.
i.e. C_i will be determine as follows:

(i) Find the robust estimate of mean 'm' and covariance matrix $'\sum'$ of the data set A.
(ii) Calculate Mohalanobis distance $A_s\ (m, p_j)$ from any point $p_j \in A$

where,

$A_s\ (m, p_j) = \sqrt{(m - p_j)^t \sum^{-1} (m - p_j)}$
in which the initial radius is
$R = {}^{Min}_{i \in n}\{A_s(m, p_j)\} + \frac{{}^{Max}_{i \in n}\{A_s(m,p_j)\} - {}^{Min}_{i \in n}\{A_s(m,p_j)\}}{(1+\varphi)^2}$

where φ is a golden ratio.

(iii) If $A_s\ (m, p_j) < R$ then $p_j \in C_i$

Step 3. Find the final shape of the cluster, that is C_i^*

i.e. C_i^* will also be determine as follows:

(i) Compute the robust estimate of mean $'m_i'$ of C_i and covariance matrix $'\sum'$
(ii) Compute the Mohalanobis distance of each point of C_i
i.e. $A_s\ (m_i, p_{ij}) = \sqrt{(m_i - p_{ij})^t \sum^{-1}(m_i - p_{ij})}$
and the radius of ith cluster given as
$R = \underset{i \in n}{Min}\{A_s(m_i, p_{ij})\} + \frac{\underset{i \in n}{Max}\{A_s(mi,p_{ij})\} - \underset{i \in n}{Min}\{A_s(m_i,p_{ij})\}}{(1+\varphi)^2}$
(iii) If $A_s\ (m_i, p_{ij}) < R_i$ then $p_j \in C_i$ else go for p_{j+1}. Repeat the process for entire data set until $j =$ size of the data.
After some iteration there will be no change in the shape of C_i, and the final shape of C_i that is C_i^*

Step 4. Formation of clusters from the remaining data set.

i.e. Once the formation of cluster C_i^* is over, then check the remaining data A - C_i^*

(i) If the remaining data is non empty i.e. A - $C_i^* \neq \phi$ go to step 2 of the algorithm and perform all the steps.
Otherwise
(ii) Stop the procedure when A - $C_i^* = \phi$.

Step 5. Data are modeled by cascade backward propagation network. i.e the cascade backward propagation neural network is fitted to the data.

Step 6. Determine the number of hidden units, i.e.

(i) Train the network with initial "K" hidden neurons in each of the hidden layer.
(ii) The value of "K" is generally taken higher than the number of input variables say "P"

Step 7. Take the computed output of the trained network and calculate the errors. i.e.

$e_t = Y_i - \hat{Y}_i$; $i = 1, 2, ..., n$. n been the size of the training set.
(i) If all the steps are satisfied by the calculated error; then select "K" as number of hidden units else
(ii) Change the number of hidden units to $K - i$, $i = 1, 2, ..., K - 1$ for the purpose of finding a network architecture that preserved the error assumption of regression model. repeat this untill the required network architecture is obtained

Step 8. Evaluate the mean squared error (MSE) use to train the cascade backward propagation neural network.

The backpropagation algorithm tends to look for the minimum of the error function in weight space using the method of gradient descent. The permutation of weights which lessens the error function is considered to be a solution of the learning problem. However the continuity and differentiability of the error function must be guaranteed since the method considered computation of the gradient of the error function at each iteration step.

4 Results and Discussion

This section discuss and analyzed the findings of this study. We start by given a brief information about the data set used in this study.

4.1 Information on data set used

The data set used in this study were obtained from the UCI machine learning repository data link [19]. Airfoil data, Aboline data and Airline data sets were selected for this study. The three (3) data sets were selected based on the fact that they contained apparent heteroscedasticity as well as sample size and number of explanatory variables. The Airfoil data, which was gathered from various tests done on two and three dimensional airfoil blades within a large wind tunnel has 1503 entries with 5 explanatory variables and 1 response variable. The Aboline data set is a well known machine learning data set that is usually used for predicting the ages of aboline using physical measurements from a large sample shell which is determined by cutting the shell through the cone, staining it and counting the number of rings through a microscope. The original data obtained from the foregoing link numbered 4177 and 8 attributes. The airline passenger data set contained the monthly airline passenger records. This data set has only two variables numbered 144 sample size. The Three data sets contained no missing values and were relatively clean, hence, applied to establish the empirical instances of our developed ($CBPNFDCARM$) model. The three data sets were then preprocessed to get ready with model variables so as to predict the performance measure based on the build model. Airfoil data, Aboline data and the Airline passenger data further fit into our algorithm with the aim of clustering each of the data sets into two clusters. The purpose of this clustering is to remove the outliers from each of the data sets before training can take place. Part of the data that contained the removed outliers i.e. (clean part) were then used in training the cascade forward back propagation neural network with the aim of determining the performance measure of the network. The mean squared error (MSE) was used as the measure of the performance of the network. The sample statistics of the data used for this study were presented in Table 1.

Table 1. Sample Statistics of the data used for the study

Data	Sample size	Mean	Variance	Standard deviation	Reference
Airfoil Data	1503	511.5012	2.7856e+06	1.6690e+03	UCI [19]
Aboline Data	4177	2.0849	3.0785e+03	55.4844	UCI [19]
Airline passenger	144	1.1176e+03	7.1074e+05	843.0546	UCI [19]

4.2 Analysis of results

The influence of non constant error variance on the performance of analysis be it regression or neural network may usually depend on factors like number of independent variables, amount of variation in variance and the size of the sample. In this study, the cascade backward propagation neural network over a filtered data by clustering algorithm based on robust measure ($CBPNFDCARM$) was considered. Predicting the performance measure hinge on Three (3) independent data set. The method of analysis in this study are weighted least square regression (WLS), cascade backward propagation neural network ($CBPN$) and cascade backwaard propagation neural network over filtered data by clustering algorithm based on robust measure $CBPNFDCARM$. The coefficient of regression in this study is estimated using the weighted least square regression and to compare the results of this methods with the neural networks that was trained using the proposed method as described in this study. The clustering algorithm that was developed for this study uses the three (3) independent data set and divide each of the data into two clusters. i.e. the part that tends to remove the outliers from the data set (clean part) and the outlier part. Robust estimate of mean and covariance matrix were employed in defining the radius of the cluster. The clean part of the data, i.e. the cluster whose elements are most as clean part were further divided into training and testing data. 70% of the data was used for training while the remaining 30% for testing the network using the back propagation learning algorithm in order to determine the measure of performance of the network. Mean squared error (MSE) was used as the performance measure of our model. The weighted least square (wls) model uses the regression analysis. In this case, weights are been added to each of the data set and the resulting result were fitted into a regression in order to find the error of regression of the weighted least squared. The prediction results of both models i.e $CBPN$, WLS and $CBPNFDCARM$ were then compared. However, in comparing the three models we tried to find the model that produces a prediction with the minimum mean squared error. Table 2 illustrates the comparison of the results. From Table (2), we evaluate the prediction performance by calculating the mean squared error (MSE) and the root mean squared error (RMSE) along with the total epoch and the total time in seconds taken by each model to converge. The results were shown in Table 2.

Table 2. Performance results of ($CBPN$), (WLS) and ($CBPNFDCARM$) for the used data set

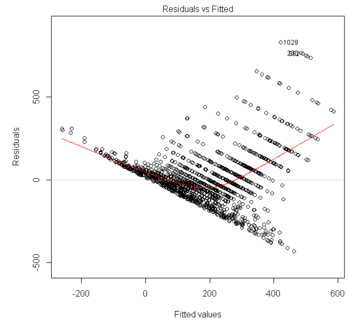
Data	Method	MSE	$RMSE$	Epoch	Time(s)
	$CBPN$	4.11e+05	641.0928	999	183
Airfoil Data	WLS	92.78	9.6322	***	***
	$CBPNFDCARM$	2.22	1.4900	1000	60
	$CBPN$	4.17e+04	204.2058	1000	1689
Aboline Data	WLS	0.09753	0.3123	***	***
	$CBPNFDCARM$	6.15e-05	0.0078	1000	1490
	$CBPN$	0.328	0.5727	527	15
Airline passenger	WLS	2.886	1.6988	***	***
	$CBPNFDCARM$	0.212	0.4604	151	10

Table (2). indicated that, the new developed method i.e. the $CBPNFDCARM$ mean squared error (MSE) performed better than weighted least squared (WLS) as well as the stand alone cascade backward propagation neural network. The mean squared errors decreased consecutively from $CBPN$ to WLS to $CBPNFDCARM$ for the airfoil and aboline data set. i.e.

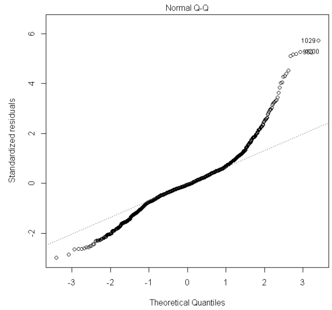
$$MSE(error)_{CBPNFDCARM} < MSE(error)_{WLS} < MSE(error)_{CBPN}$$

Because of the iterative nature of neural networks, it usually takes more time to converge than regression analysis. However, when functional form of heteroscedasticity is not known, our proposed method seems to be a better alternative. From Table 2, even though both models have same total epochs in some cases, the time it takes for the developed model to converge is less than the time taken for the cascade backward propagation neural network to converge. This is an assertion that the developed model converges faster than the standalone cascade backward propagation neural network model which further illustrate that our developed model outperform the stand alone model. Residuals vs fitted as well as the Q-Q plots from weighted least square (WLS) of each data used in this study were shown in Fig. 2.

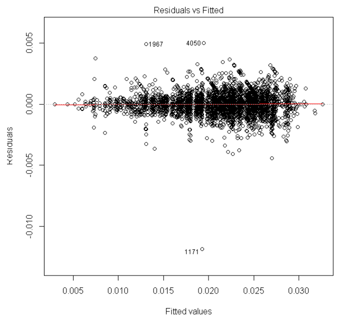
Weighted least square been an estimate procedure which is usually followed when errors in a regression models are uncorrelated and do not have equal variances were used in determining the residuals and the Q-Q normality plot for all the data set used. Fig. 2 showed the (wls) residual vs fitted and Q-Q normality plots of Airfoil, Aboline and Airline passenger data sets. From the plots of the (wls) residual vs fitted of all the data sets, it has shown the effect of weighted least square as an estimation procedure in reducing the presence of heteroscedasticity in data. For the neural network architecture, the performance of our developed model ($CBPNFDCARM$) and the ($CBPN$) were plotted. The plots of the performance were shown in Fig. 3.



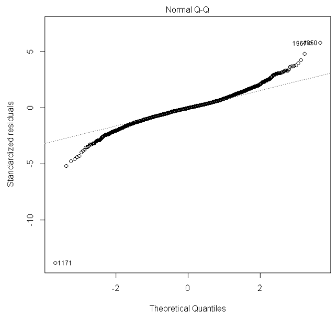
(a) Residual plot for Airfoil data



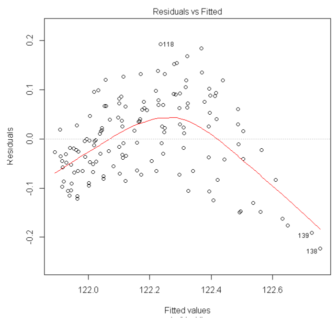
(b) Q-Q plot for Airfoil data



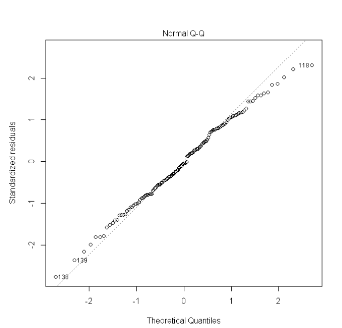
(c) Residual plot for Aboline data



(d) Q-Q plot for Aboline data

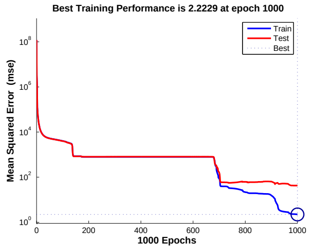


(e) Residual plot for Airline passenger data

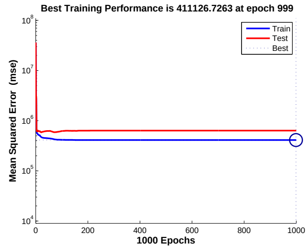


(f) Q-Q plot for Airline passenger data

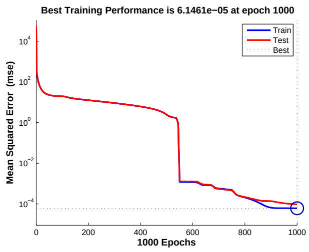
Fig. 2. WLS **Q-Q normality and Residual plot for Airfoil, Aboline and Airline passenger data**



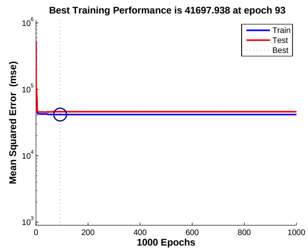
(a) Performance plot for ($CBPNFDCARM$) of Airfoil data



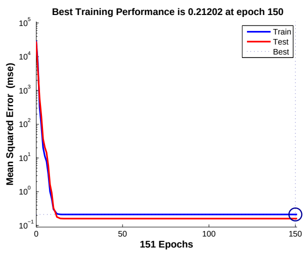
(b) Performance plot for ($CBPN$) of Airfoil data



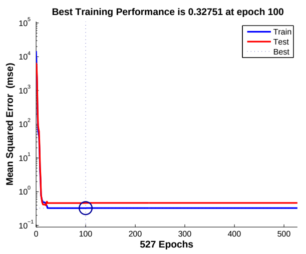
(c) Performance plot for ($CBPNFDCARM$) of Aboline data



(d) Performance plot for ($CBPN$) of Aboline data



(e) Performance plot for ($CBPNFDCARM$) of Airline passenger data



(f) Performance plot for ($CBPN$) of Airline passenger data

Fig. 3. Performance plot for ($CBPNFDCARM$) and ($CBPN$) of Airfoil Aboline, and Airline data

Fig. (3a) showed the performance plot of $CBPNFDCARM$ for Airfoil data set. From the plot, the mean square error for best training performance at epoch 1000 was 2.2229 which was far better than the mean square error for the best training performance of the $CBPN$ of the same Airfoil data of 4.11e+05 as shown in Fig. (3b). Also, the mean square error best training performance of the Aboline data for $CBPNFDCARM$ in Fig. (3c) at epoch 1000 was 6.15e-05, while for $CBPN$ in Fig. (3d) of the same Aboline data was 4.17e+04. Again, in Fig. (3e), the mean square error for best training performance of $CBPNFDCARM$ for the Airline passenger data was 0.212 compared to the mean square error of $CBPN$ of the same Airline passenger data which stands as 0.328 as can be seen in Fig. (3f). The plots of the figures from all the data sets further illustrate that, our developed model ($CBPNFDCARM$) has a better performance measure compared to the standalone cascade backward propagation neural network ($CBPN$).

5 Conclusions

In this study, we have developed and propose a new advanced version of backward propagation neural network based clustering technique. Experiments comparison between this approach (based on clustering technique) and least squared technique were made. The new developed approach is better within in heteroscedasticity given evaluation metrics MSE. The main point of this research is to use clustering method to prune training data, which is the most significant invention. For further research, we try to feed clean part of the data into regression technique. If regression model is applied to non-outlier data, its efficiency may be enhanced. In general, linear regression model cannot be better than neural network when neural network supports inside non-linear mechanism but linear regression model is simpler. However this research is very significant because it improves the traditional backward neural network. This work being empirical as well as limited to some specific functional forms. Therefore, we suggest further work to be carried out to strengthen the findings of this work.

Acknowledgement

The research is partly financed by FRGS grant (203/PMATHS/6711368) from the ministry of Higher Education

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Hadi AS. Identifying multiple outliers in multivariate data. Journal of the Royal Statistical Society. Series B (Methodological). 1992;761-771.

[2] Hadi AS. A modification of a method for the detection of outliers in multivariate samples. Journal of the Royal Statistical Society. Series B (Methodological). 1994;393-396.

[3] Nag AK, et al. Multiple outlier detection in multivariate data using self-organizing maps title. Computational statistics. 2005;20(2):245-264.

[4] Alih E, Ong HC. Robust cluster-based multivariate outlier diagnostics and parameter estimation in regression analysis. Communications in Statistics-Simulation and Computation; 2014. (in press).

[5] Carroll RJ, Ruppert D. Transformation and weighting in regression. CRC Press.1988;30.

[6] Daye ZJ, et al. High-dimensional heteroscedastic regression with an application to eQTL data analysis. Biometrics. 2012;68(1):316-326.

[7] Brem RB, Kruglyak L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. In: Proceedings of the National Academy of Sciences of the United States of America. 2005;102(5):1572-1577.

[8] Lim C, et al. Statistical inference in nonlinear regression under heteroscedasticity. Sankhya B. 2010;72(2):202-218.

[9] Montgomery DC, et al. Introduction to linear regression analysis. John Wiley & Sons; 2015.

[10] Gujarati DN. Basic econometrics. Tata McGraw-Hill Education; 2009.

[11] Ong HC, Alih E. A control chart based on cluster-regression adjustment for retrospective monitoring of individual characteristics. PloS one. 2015;10(4):e0125835.

[12] Pwasong A, Sathasivam S. A New hybrid quadratic regression and cascade forward backpropagation neural network. Neurocomputing; 2015.

[13] Paliwal M, Kumar UA. The predictive accuracy of feed forward neural networks and multiple regression in the case of heteroscedastic data. Applied Soft Computing. 2011;11(4):3859-3869.

[14] Thatoi D, et al. Comparison of CFBP, FFBP, and RBF networks in the field of crack detection. Modelling and Simulation in Engineering. 2014;2014:3.

[15] Singh R, Srivastava L. Line flow contingency selection and ranking using cascade neural network. Neurocomputing. 2007;70(16):2645-2650.

[16] Hedayat A, et al. Estimation of research reactor core parameters using cascade feed forward artificial neural networks. Progress in Nuclear Energy. 2009;51(6):709-718.

[17] Rana M, et al. A robust modification of the goldfeld-quandt test for the detection of heteroscedasticity in the presence of outliers. Journal of mathematics and Statistics. 2008;4(4):277-283.

[18] Atkinson A, Riani M. Robust diagnostic regression analysis. Springer Science & Business Media; 2012.

[19] Bache K, Lichman M. UCI machine learning repository; 2013.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/15994