FORECASTING PERFORMANCE OF CASCADE FORWARD BACK PROPAGATION NEURAL NETWORK FOR DATA WITH OUTLIERS

MAMMAN MAMUDA

UNIVERSITI SAINS MALAYSIA

2017

FORECASTING PERFORMANCE OF CASCADE FORWARD BACK PROPAGATION NEURAL NETWORK FOR DATA WITH OUTLIERS

by

MAMMAN MAMUDA

**Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy**

March 2017

ACKNOWLEDGEMENT

Alhamdulillahi Rabbil′ Alamin.

I wish to express my sincere gratitude to my supervisor, Associate Professor Saratha Sathasivam for her constant support, guidance, continuous encouragement, patience and constructive ideas throughout my research work. I have learned so much from her. I also express my sincere appreciation to all members of staff in the School of Mathematical Sciences Universiti Sains Malaysia (USM) Pulau Pinang for their support and encouragement for the whole period of study.

My special thanks and gratitude goes to the management of Niger State College of Education Minna (COE MINNA) for their financial and moral support throughout my period of study. Warmest thanks to my fellow postgraduate students in the school of Mathematical Sciences, USM and to my country men as well as women studying in USM for their constant support and encouragement.

I wish to express my utmost thanks to my wives, my mother, brothers, sisters, children and siblings for their prayers, love, patience and advice during the whole period of my study. Their continuous support and encouragement are highly appreciated. I also want to acknowledge my late father who passed away during the period of my study and whose last words and wishes have been a source of strength to my academic pursuit. May your soul continue to rest in peace. Amin thumma Amin.

Finally, and most importantly, I would like to thank the Federal Government of my

dear country Nigeria for making funds available through the tertiary education trust fund (TETFUND) to finance my PhD program at the Universiti Sains Malaysia, Pulau Pinang. May God continue to bless Nigeria, Amin.

Indeed no man is an Island on his own.

Mamuda Mamman

2017

TABLE OF CONTENTS

CHAPTER 1 **–INTRODUCTION**

CHAPTER 2 **– LITERATURE REVIEW**

CHAPTER 3 **–THE EXISTING MODELS**

CHAPTER 4 **–THE PROPOSED** $CFBNFDCARM$ **MODEL**

LIST OF PUBLICATIONS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
ARX	Autoregressive with Exogenous
BCM	Bienenstock-Cooper-Munro
BDP	Break Down Point
BP	Backpropagation
BPNN	Backpropagation Neural Network
CFBNN	Cacade Forward Backpropagation Neural Network
CFBNFDCARM	Cascade Forward Backpropagation Neural Network Filtered Data by Clustering Algorithm based on Robust Measure
CI	Confidence Interval
COA	Cuckoo Optimization Algorithm
CSR	Cascade Shape Regression
CV	Contribution Value
DetMCD	Deterministic Minimum Covariance Determinant

DRMAT	Dimension Reduction-based Model Adaptive Test
ERP	Earth Rotation Parameters
ERNN	Elman Recurrent Neural Network
ESN	Echo State Network
FFBN	Feed Forward Backpropagation Neural Network
FNN	Feed Forward Neural Network
FSA	Forward Search Algorithm
GA	Genetic Algorithm
GLM	Generalized Linear Model
GM	Geometric Mean
IRWLS	Iterative Regression Weighted Least Square
IWLSSVR	Iteratively Weighted Least Square Support Vector Regression
LM	Linear Model
LMS	Least Median Square
LRNN	Layered Recurrent Neural Network
LS	Least Square

LS-AR	Least Square Autoregressive
LSSVR	Least Square Support Vector Regression
LTD	Least Trimmed Difference
LTS	Least Trimmed Square
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MCD	Minimum Covariance Determinate
MLP	Multi-Layer Perceptron
MLR	Multiple Linear Regression
MMD	Minimum Mahanalobis Distance
MPG	Miles per Gallon
MSE	Mean Square Error
MVE	Minimum Volume Ellipsoid
NASA	National Aeronautics and Space Administration
NARX	Nonlinear Auto-Regressive with Exogenous
NEMMCO	National Electricity Markets Management Company

NN	Neural Network
OLS	Ordinary Least Square
OME	Ordinary M-Estimation
PAH	Polycyclic Aromatic Hydrocarbon
PTE	Preliminary Test Estimation
QAP	Quadratic Assignment Procedure
Q-Q	Quantile-Quantile
QRM	Quadratic Regression Model
RGM	Reweighted Geometric Mean
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SCNN	Self-Organized Cascade Neural Network
SSE	Sum Square Error
TWC	Two Way Clustering
UCI	University of California Irvine
USM	Universiti Sains Malaysia

WLS	Weighted Least Square
WLS-AR	Weighted Least Squared Autoregressive
WME	Weighted M-Estimation

LIST OF SYMBOLS

β	Regression parameters
β_j	Element of β related to the j^{th} regressor
b_w	Scale parameter
C^{-1}	Covariance matrix
D_{max}	Maximum time complexity of a single model
Δw_{ij}	Weight updating function
δ_j	Negative of the gradient of the output error function
d_i^2	Ordered sequence of the distance in the minimum Mahanalobis distance
ε_i	Random error at the i^{th} response variable
ε_t	Random error at time t
e_t	Error term in regression model
e_i	i^{th} least square residual
f	Sigmoid function
$g_{(x)}$	Cascade Forward Backpropagation Neural Network

J	Jacobian matrix
J^T	Transpose of Jacobian matrix
η	Learning rate
m	Dataset in minimum Mahalanobis distance
μ	Damping factor
$o(n)$	Time complexity
Ω	Estimate of variance
∂	Partial derivative
Φ	Shape-Index feature descriptor
P_j	Mean of the basic subsample
φ	Connection weight connecting output layer neuron to hidden layer neuron
R	Correlation coefficient
R^2	Coefficient of Determination
S_w^2	Weighted Squared Residual
σ^2	Variance
Σ	Summation

T	Asymptotic or centripetal time complexity
$V(\hat{\beta})$	Covariance matrix with heteroscedasticity
$\forall$	For all
w	Weights
w_j	$m \times 1$ Weight Vector
w_0	Bias Term
W^T	Linear transform matrix
X	Independent variable
X_i^*	Shape of the face of a Cascade Linear Regression
X_i^{T-1}	Initial shape of the face of a Cascade Linear Regression
x_i	$m \times 1$ Input vector
x_{ji}	p-Predictors variables
Y	Dependent variable
$\hat{Y}$	Minimum Mahalanobis distance
Y_i	Preferred output
$\hat{Y}_i$	Predicted output

y_{ki}	q-response variables
Z_y	Given dataset
$Z_y^{'}$	Subset of a given dataset

PRESTASI PERAMALAN RANGKAIAN NEURAL RAMBATAN BALIK LATA HADAPAN BAGI DATA TERPENCIL

ABSTRAK

Dalam kajian ini, satu rangkaian neural berasaskan pengelompokan telah dibangunkan untuk menyiasat dan membandingkan prestasinya dengan teknik-teknik pemodelan lain bagi kes penyimpangan andaian berkaitan hubungan homoskedastik dalam set data. Enam set data dengan saiz sampel yang berbeza telah dimodelkan menggunakan lima teknik pemodelan. Kelima-lima model tersebut ialah rangkaian neural berasaskan pengelompokan yang dirujuk sebagai rangkaian neural rambatan balik lata hadapan atas data yang telah disaring menggunakan algoritma pengelompokan berdasarkan ukuran teguh (CFBNFDCARM), rangkaian neural rambatan balik lata hadapan kendiri (CFBNN), rangkaian neural suapan hadapan kendiri (FFNN), rangkaian neural berulang Elman kendiri (ERNN) dan teknik regresi kuasa dua terkecil berwajaran (WLS). Keenam-enam set data telah diplot menggunakan plot kotak bagi mengenalpasti kehadiran data terpencil dalam set data. Teknik yang dicadangkan iaitu CFBNFDCARM mempunyai tiga ciri berlainan yang membezakannya dengan sebarang rangkaian neural berasaskan pengelompokan sedia ada yang diperolehi daripada literatur terutamanya bagi kes kehadiran data terpencil dalam set data. Pertamanya, teknik pengelompokan telah digunakan ke atas setiap set data untuk menyaring data terpencil kerana kehadirannya dalam data membawa kepada penyimpangan andaian kehomoskedastikan pada sesuatu model. Kehhadiran data terpencil menyebabkan ketaksamaan varians dalam model yang membawa kepada hubungan heteroskedastik.

Keduanya, algoritma pengelompokan membahagikan data kepada dua bahagian, iaitu bahagian yang data terpencilnya telah dikeluarkan(bahagian bersih) dan bahagian yang mengandungi data terpencil. Ketiganya, bahagian bersih telah dibahagikan kepada set data latihan dan set data ujian dan dipadankan ke dalam rangkaian neural bagi menentukan ukuran prestasi rangkaian tersebut menggunakan algoritma pembelajaran penurunan kecerunan. Sebanyak 70% daripada data telah digunakan bagi melatih rangkaian manakala bakinya telah digunakan bagi menguji rangkaian tersebut. Teknik pengelompokan yang dicadangkan telah menggunakan jarak Mahalanobis minimum terhadap anggaran lokasi dan penyebaran yang menggunakan anggaran teguh min dan kovarians matriks dalam mentakrifkan jejari bagi algoritma pengelompokan. Keputusan yang diperolehi daripada rangkaian neural berasaskan teknik pengelompokan yang dicadangkan (CFBNFDCARM) telah dibandingkan dengan keputusan yang diperolehi daripada CFBNN, FFNN, ERNN dan WLS. Perbandingan tersebut menunjukkan bahawa, secara puratanya, prestasi peramalan teknik CFBNDCARM telah mengatasi prestasi peramalan teknik-teknik kendiri (CFBNN, FFNN, ERNN) dan WLS berdasarkan penilaian metrik min ralat kuasa dua, punca min ralat kuasa dua, jumlah kuasa dua ralat, min ralat mutlak dan min peratusan ralat mutlak.

FORECASTING PERFORMANCE OF CASCADE FORWARD BACK PROPAGATION NEURAL NETWORK FOR DATA WITH OUTLIERS

ABSTRACT

In this research, a clustering based neural network was developed with the aim of investigating and comparing its performance with the performance of other model techniques in the case of deviation from the assumption of homoscedastic relationship in dataset. Six dataset with different sample size was modeled using five model techniques. The five models are the clustering based neural network that is refers to as the cascade forward backpropagation neural network over a filtered data by clustering algorithm based on robust measure ($CFBNFDCARM$), the standalone cascade forward backpropagation neural network ($CFBNN$), the standalone feed forward neural network ($FFNN$), the standalone Elman recurrent neural network ($ERNN$) and the weighted least square (WLS) regression techniques. Each of the six dataset were plotted using box plot in order to identify the presence of outliers in the dataset. The proposed $CFBNFDCARM$ technique possesses three distinct features that differentiate it from any existing clustering based neural network techniques obtainable in the literature especially in the case of outliers in dataset. Firstly, the clustering technique was employed to each of the data set to filter out the outliers since they lead to deviation of a model from the assumption of homoscedasticity. Presence of outliers in dataset leads to unequal variance in a model which inturns result to a heteroscedastic relationship. Secondly, the clustering algorithm divides the data into two parts, a part that contained the removed outliers (clean part) and a part that contained the outliers. Thirdly, the clean

part of the data were further divided into training and testing dataset and fitted into the neural network in order to determine the measure of the performance of the network using the gradient descent learning algorithm. 70% of the data was used for training the network while the remaining 30% was used for testing the network. The proposed clustering technique used the minimum Mahalanobis distance of estimation for location and dispersion that employed the robust estimate of mean and covariance matrix in defining the radius of the clustering algorithm. The results obtained from the proposed clustering based neural network $CFBNFDCARM$ technique were compared with the results obtained from $CFBNN$, $FFNN$, $ERNN$ and WLS. The comparison indicates that the emerging forecasting performance results from the proposed $CFBNFDCARM$ technique generally on the average outperformed the forecasting performance results from the standalone $(CFBNN, FFNN, ERNN)$ and the WLS regression techniques in terms of the evaluating metrics of the mean square error, root mean square error, sum square error, mean absolute error and mean absolute percentage error.

CHAPTER 1

INTRODUCTION

1.1 Overview

Regression analysis have dominated the problems of function estimations that has to do with finding the probabilistic relationship between the dependent variable Y and the set of independent variables X described by the conditional distribution $Y \mid X$ from a given regression model $Y = f(X)$. Significant interest in alternative methods has suffices in the last few decades. One reason for the interest is the increase of awareness and sensitivity to the problems that occur with the constant application of regression analysis. In regression modeling, experimental data that are usually used may assume a linear pattern, but contaminated or spurious data points may be inevitable. Presence of contamination or outliers in data set seem to be very common and as such leads to poor estimation which in turn not a true representation of the model parameters. The over increase of size and complexity of data over the last decade have encourage the development of different and sophisticated methods in dealing with size and complex data sources. Modeling using real data as asserted by Hampel et al. (2011) are presumed to contains between 1% to 10% contamination. In modeling complex system of linear and non-linear structures, there is the need of developing a robust method constituting an effective fashion to improve the assumption unequal variance in a model which inturns result to a heteroscedastic relation. Outliers in regression analysis leads to deviation of a model from the assumption of homoscedasticity. One of the classical regression theory that deals with this problem is the weighted least square (WLS)

method. The discovery of personal computer in the 1970s has given a breakthrough in handling procedures and approximation algorithms that were once seen as too complex to compute. Geometric growing rate of applications that involve computer technology has become almost trivial for computational confidence developed to emphatically pursue new approaches and new algorithms.

Over the last decades, artificial neural network (*ANN*) methods, regression methods and heteroscedasticity are vibrant areas of research of the present, the future as well as the last few decades as far as estimation and approximation procedures are concern. In various decision making theories such as in medicine, economics, business, industries, education and engineering, the strictness of *ANN* method, regression methods and heteroscedasticity is found to be very essential. In improving the usefulness of existing estimation and approximation models as well as to develop an efficient models for estimation and approximation, numerous investigations are ongoing. Recently, *ANN* have become one of the most regular methods in exploration of estimation and approximation practice within their nonlinear modeling ability. Also, regression analysis in the presence of outliers have displayed their capacity in evaluating the impact of their predictive ability (Payne, 2014). However, the inability of standalone *ANN* methods and regression analysis to produce absolute accuracy in the presence of outliers in data set is of great concern, they are not generally accepted as the widely best models that can be used in virtually every estimation and approximation circumstances in outlier's dataset even when they posses the capacity of modeling and estimating a various range of issues and events.

ANN which is seen as an information processing system that is designed to model the capability of biological neurons of human like brain as well as a known technique with the ability to estimate functional relationship. Biological neural structures consist of neurons, which constitute a network through synaptic connections in a mix of axons and dendrites. A chemo-electrical motions converses the neurons with these connections to establish behavior. Lippmann (1987) defined artificial neural network (*ANN*) as the network that consist of interrelated neurons that maneuver in parallel and linked collectively through weights. Carmichael (2001) described neural network as a link that comprises of a quantity of allied nodes. Related to each of the link are the numeric weight, the weight are seen to be the key means of long-term storage in the network where learning takes place by updating the weights. Definite nodes are linked to the outside setting, which are usually taken as input or output nodes. Ibrahim et al. (2009) asserted that *ANN* have been applied in various submissions such as pattern classification, image and signal processing and several other intelligent systems. Backpropagation neural network (*BPNN*) is seen to be an important method in *ANN* which is a multi-layer feed-forward artificial neural network that is very useful whenever the network design is appropriately taken. The learning algorithm used in training the *BPNN* is the backpropagation learning algorithm, but having a very slow convergence rate as its setback. Lippmann (1987) asserted that learning the problem in an extremely minor network is not sufficient, but over fitting and unfortunate generalization of performance will transit in an extremely huge magnitude. The development of speed of the convergence of network have been supported by various techniques. Many literatures as asserted by (AL-Allaf, 2012) deliberated the use of various *ANN* architectures and training algorithms for image density in advancing the speed of convergence to deliver high density ratio. Today, the existence of unknown and potential nonlinear relationships between the response and explanatory variables are dealt with using modern regression methods. Independently, researches on new methods using basic linear regression models in handling the problems of heteroscedasticity, non-normality and outliers were made. In classical regression models, homoscedasticity, assumption of equal variance is very fundamental (Carroll and Ruppert, 1988). Wilcox and Keselman (2012) declared that even though the assumption of equal variance are theoretically convenient, it is often not satisfied within practice.

Bingyun and Malin (2009) asserted that regression models has the skilled to analyze the resolute correlation between response variables and predictor variables, which in turns predict the style of the predictor variables. A model of regression possesses a resolute firmness; and it is from time to time demanding to collect information and predict a unique creation or a novel pattern that is up and coming in resolving the demand informations. Consequently, new predicting techniques using the integration of regression and neural network models which are established in the event where the assumption of equal variance (homoscedasticity) is deviated and set up a heteroscedasticity relationship in data sets are merged in order to be able to overcome the challenges characterized by the existence of unequal variance in data sets. Warner and Misra (1996) declared that the neural network in the field of prediction/estimation problems contribute significantly in putting the field on a strong theoretical and conceptual foundation. Paliwal and Kumar (2011) entrenched a hybrid model with the integration of regression and neural network techniques and employed it in comparing the performance of the neural network and the regression technique to prognosticate the assumption of deviation from homoscedastic relationship. Obviously, the WLS regression technique and the feed forward neural network $(FFNN)$ technique used by them is a very credible model to employ in predicting/estimating models in heteroscedastic data sets. However, the reliability of prediction/estimation obtained from their techniques are not very stable especially when the data sets contain many form of heteroscedasticity, and does not convey sufficient details to a decision maker. Therefore, their regression technique integrated with a neural network model, will exhibit its distinctive performance and hold back its deficiency.

1.2 Modeling with Linear and Nonlinear Models

Models that constitute a constant or product of a parameter and a predictor variables with one basic form are refers to as linear models. A nonlinear model is a model that is unlike the linear models takes many different forms. The estimating equations in a nonlinear model whose solution yields the parameter estimates depend on the fashion of nonlinear parameters. The estimation problem of nonlinear model typically have no closed form solution, hence, can only be solved by iterative, numerical technique. This study is hinged on modeling of a linear method and four (4) nonlinear methods. The linear method to be focus in this study is the WLS regression method while the four (4) nonlinear method focus here are the $CFBNFDCARM$, the standalone $CFBNN$, $FFNN$ and $ERNN$ methods.

In this thesis, $CFBNFDCARM$ method tends to serve as our singular contribution to the nonlinear modeling in clustering based neural network practice.

The $CFBNFDCARM$ possessed distinct features that differentiate it from any current clustering based neural network performance measures obtainable in the literature. Firstly, by removing the outliers from a given data set using the clustering algorithm, the clustering algorithm tends to divide the data set into two part, a part that contains the removed outliers $(cleanpart)$ and the part that contains the outliers $(outlierpart)$

(Atkinson and Riani, 2012). Secondly, the clean part of the data set i.e. the removed outlier part were then fitted into the neural network for training and testing using the gradient descent learning algorithm as in (AL-Allaf, 2012) with the aim of determine the measure of performance of the network.

The mean squared error (MSE), root mean squared error $(RMSE)$, mean absolute error (MAE), sum squared error (SSE) and the mean absolute percentage error $(MAPE)$ were the performance measure employed to evaluate the performance of the developed as well as the standalone models in this thesis. These performances were then collated and extensively deliberated upon with other linear and nonlinear models mentioned in this study.

1.3 Regression and Neural Network Methods

Regression study is a dynamic research area in which researchers are absorbed into the most recent decade(Paliwal and Kumar, 2011). Regression study can be seen as the study of how two sets of variables are related to each other. The first set of the variable are sometimes called the regressor or independent variables defines the p-predictors, where p defines the number of the independent variables usually denoted as:

$x_{ji} = x_{1i}, x_{2i}, x_{3i}, ..., x_{pi}; i = 1, ..., n; j = 1, ..., p.$

These predictor variables as assumed by Gaussian paradigm are fixed. The second set of variable which are sometimes called the q-response or the dependent variable, where q defines the number of the dependent variables denoted as:

$y_{ki} = y_{1i}, y_{2i}, ..., y_{qi}; i = 1, ..., n; k = 1, ..., q.$

They are often called dependent variable because their value depend on the regressor variables. Neural networks are mathematical tools of both linear and nonlinear functions that has the capability of human like brain with the potentiality of predicting problems (Gorr et al., 1994; Warner and Misra, 1996). The functional potentiality of neural network has led to a verse studies comparing its performance in the predictive capability and regression analysis. Jiang and Chen (2016) asserted that, even though *ANN* are good in predicting and assumption models, they however have a very weak convergence rate and sometimes traps to a local minimum which is found to be its setback. Research on using different techniques and regression methods have been proposed in literature by Zhangzhen and Tianhe (2012), Behnasr and Jazayeri-Rad (2015) and Guo et al. (2014).

Recently, regression with neural networks (NNs) models has gained ground in such a way that many researchers in the field of prediction and estimations modeling are being draw into these predictions and estimations methodologies. There are numerous *ANN* modeling and regression methods in literature. The frequently used *ANN* method in literature is a single hidden layer $(FFNN)$ known as the multi-layer perceptrons $(MLPs)$ (Zhang et al., 1998; Zhang, 2003). The singular characteristic of *ANNs*, when applied to modeling and prediction/estimation methods is their integral proficiency in modeling nonlinear functions. This deserts any assumption on the statistical distribution possessed by regression analysis. Chithra et al. (2016) proposed an *ANN* technique to compared its predicting performance of concrete containing nano silica and copper slag with the precision of multiple regression analysis. The performance of neural network and the generalized regression was proposed by del Rosario Martinez-Blanco et al. (2016) with the aim of comparing the performance of *BPNN* and the generalized regression neural network in a solution of neutron spectrometry problems. The applicability in finding complex functional relationship is one of the important aspect of neural networks (Thatoi et al., 2014). Unlike the generalized linear models, it is not necessary to pre-specify the type of relationship between covariant and response variables. This fact makes neural networks a valuable statistical tool which are in particular direct extensions of generalized linear models and can be applied in a similar manner. Various researchers in the area of prediction and estimation using the fusion of *ANN* and regression models will agree with one that there exist numerous other current *ANN* composition in literature due to ceaseless trial studies in *ANN* and regression methodologies.

1.4 Problem Statement

Identification of outliers which is often refers to as influential observation in data sets is a common goal of a data analyst. The best estimates of observations when assumptions of Gaussian pattern are valid is usually obtained using the classical least squares (LS) regression, but it lack the sensitivity to outliers, nonlinearity heteroscedasticity as well as connectivity between the input and output variables of a given function. The presence of the influence of outliers falsify these estimates which in turn make their values to be no longer reliable. Hadi (1992) presented an algorithm for detecting outliers in multivariate data set using a robust estimate. Alih and Ong (2017) proposed the use of minimum Mahanalobis distance (MMD) to construct a cluster phase to a bounded influence regression phase with the use of robust regression method of identifying outliers. From the above assertions, it turns out that the computation of robust Mahanalobis distances from the robust location and scatter matrix forms the basis of outliers identification. These phenomena motivated this study. The new algorithm developed in this thesis will address the problems of heteroscedasticity in data sets using a robust estimates of location and dispersion matrix that helps in preserving the error assumption of linear regression.

The proposed technique will overcome some of the challenges of overfitting and overtraining that are associated with the standalone neural network techniques, since contaminated data yields to poor fitting of the overall data and form large residuals that are characterized by poor efficiency and lack of reproducibility. This research is therefore motivated by this ill-effects to detect outliers and propose an alternative clustering based neural network approach to address the inherent issue with existing alternatives.

1.5 Methodology

The homogeneity of residual variance in ordinary least square (OLS) regression allows an easy computation and forms 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 work of (Carroll and Ruppert, 1988; Rana et al., 2008) . In classical regression theory, WLS technique is one of the techniques used in dealing with present of outliers in a data set. The existence of outliers in a data set are presumed to make a model to deviate from the assumption of homoscedasticity.

Paliwal and Kumar (2011) showed that the errors obtained in training neural network are generally smaller than errors obtained using the WLS regression with more differences from the smaller sample size and becoming less for large sample size. Findings from their study further revealed that based on the fact of the capability of neural network in estimating functional relationship, $FFNN$ outperformed the WLS method.

However, the presence of outliers in data set was ignored in their study. This may contradict the assumption of normality or even both normality and homoscedasticity. Atkinson and Riani (2012) also asserted that most outlier detection techniques tends to divide the data into two parts, a part that contains the removed outliers (clean part) and the part that contains the outliers (outlier part). The clean part are then used for parameter estimations. In another development, Alih and Ong (2017) proposed a robust cluster-based multivariate outlier diagnostic in regression analysis to estimate parameters. In their proposed robust regression method, *MMD* was considered in constructing a cluster phase to a bounded influence regression phase in identifying outliers. Findings from their method shows that, the resulting proposed method has a greater advantage over other robust regression techniques. Their approach was based on regression analysis.

In this research, the use of the *MMD* will be extended by incorporating neural network approach in the case of outliers in data set. The clustering based neural network model considered in this thesis will be formulated using the *MMD* and *ANN* model. These two models will be combine to form a clustering based neural network technique. The robust estimate of mean and covariance matrix will be employed to define the radius of the clustering algorithm in this research. Our method follows the approach of Paliwal and Kumar (2011), Atkinson and Riani (2012) and Alih and Ong (2017) with more emphasis on firstly filtering the outliers from the data set using the developed robust clustering based algorithm. Six (6) different data set that were obtained from the *UCI* machine learning repository data set (Asuncion and Newman, 2007) and R data set were employed in the developed clustering based algorithm. Box plot was used to check the presence of outliers in the used dataset and the variance, skewness as well as the kurtosis of each of the used dataset was also determined. Once the outliers were removed from each of the data set, $CFBNN$ are fitted to the remaining data set that were assumed to contain no outliers for the purpose of training the network with the aid of Bayesian regularization training algorithm. The novel neural network technique to be produced in this thesis as a result of the amalgamation of the MMD and the ANN method will be called " cascade forward backpropagation neural network over a filtered data by clustering algorithm based on robust measure" $(CFBNFDCARM)$. Five (5) measure of evaluating metrics of MSE, $RMSE$, SSE, MAE and $MAPE$ were used as the performance function of the network. The obtained dataset were also introduced to the standalone neural network techniques $(CFBNN$, $FFNN$ and $ERNN)$ for training using the same Bayesian regularization training algorithm as well as the performance functions. For the WLS regression analysis, each of the obtained dataset were weighted before applying the regression analysis. The five measure of the evaluating metrics used in this thesis were also applied to the WLS regression technique. The measure of errors along with the number of epoch as well as the time taken obtained for each of the standalone neural network and the weighted least square regression were then compared with the errors obtained for the proposed neural network clustering based technique with the aim of determining the effectiveness of the developed technique. The WLS method is the technique of regression model that will be employed in this thesis, while the standalone existing $CFBNN$, $FFNN$ as well as the $ERNN$ methods will be the technique to be employed for the artificial neural network models. In addition, residuals produced as a result of fitting the weighted least square regression technique using the data sets will be entertained in this thesis. Computational experiment were conducted using $MATLAB-R2014a$ and $Ri386version3.3.0$ softwares. The method employed in this thesis is illustrated in Figure 1.1

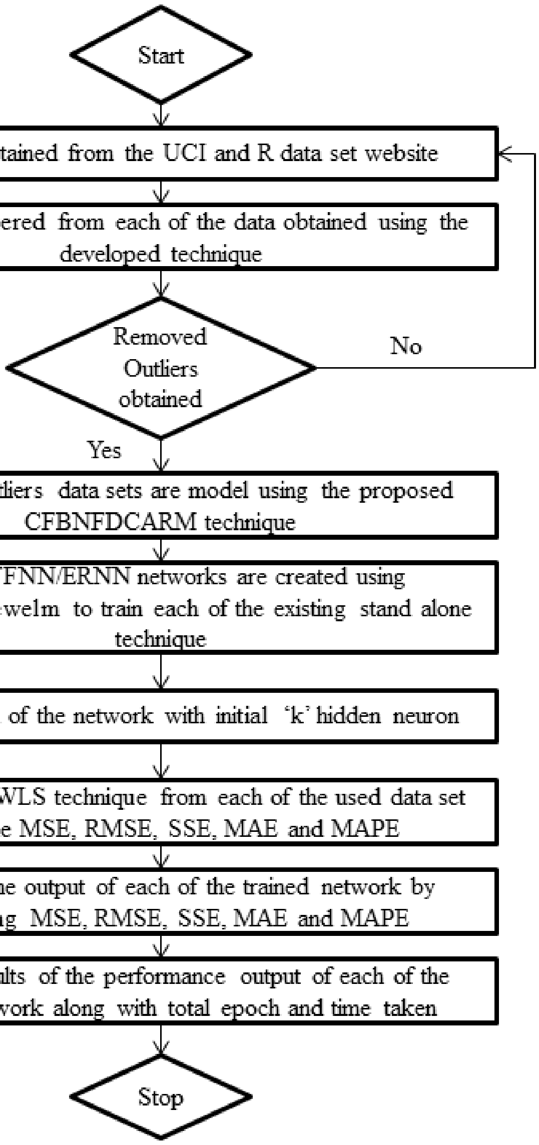
Start
Data are obtained from the UCI and R data set website
Outliers are filtered from each of the data obtained using the developed technique
Removed Outliers obtained
No
Yes
Removed outliers data sets are model using the proposed CFBNFDCARM technique
CFBN/FFNN/ERNN networks are created using newcf/newff/newelm to train each of the existing stand alone technique
Training each of the network with initial 'k' hidden neuron
Compute the WLS technique from each of the used data set using the MSE, RMSE, SSE, MAE and MAPE
Compute the output of each of the trained network by calculating MSE, RMSE, SSE, MAE and MAPE
Save the results of the performance output of each of the trained network along with total epoch and time taken
Stop

Figure 1.1: Flowchart of the methodology

1.6 Research Objectives

The main objective of this research is to develop an enhanced model of neural network through the incorporation of a robust clustering based technique. Three different aspects will be investigated such as follows:

1. Develop a new neural network model called *CFBNFDCARM* in the presence of outliers in data set.

2. Investigate and depicts how outliers in data sets affect the performance of results in modeling process.

3. Collect and compare the performance of the *CFBNFDCARM* technique with the performances of the standalone *CFBNN*, *FFNN*, *ERNN* techniques and *WLS* regression technique via the evaluating metrics of *MSE*, *RMSE*, *SSE*, *MAE* and *MAPE* in the case of outliers in data set.

1.7 Scope and Limitations of the Study

This study is confined to the problems associated with the deviation of models from the assumption of homoscedastic relationship which in turn leads to a heteroscedastic relationship. The scope of this thesis emphasized only on the presence of outliers in dataset which is one of the properties that leads to heteroscedastic relationship in models. Clustering algorithm based on robust measure is incorporated in detecting/filtering outliers that are present in the dataset employed in this study. However, the study is limited to the use of *MMD* in the clustering technique.

1.8 Thesis Organization

Chapter 1 discusses the background of predicting/estimating models of linear and non-linear methods presented in Section 1.2 and section 1.3 along side a classical background of regression and neural network methods. Chapter 2 presents a review of related literature and background introduction of linear and non-linear techniques as well as literature on regression with outliers. Appraisal literature of non-linear models in neural network as well as regression in cascade forward neural network are also presented in Chapter 2. Review on estimators for location and dispersion in the presence of outliers as well as review on robust estimators for multiple regression are also presented in Chapter 2. Chapter 3 presents a brief explanation on the existing models consider in this thesis. The developed *CFBNFDCARM* is presented in Chapter 4. A brief literature review on *MMD* estimator is also presented in Chapter 4. The criteria of *MSE*, *RMSE*, *MAE*, *SSE* and the *MAPE* are supported in this chapter. The disintegration of data, preprocessing of data as well as application of the data extracted that were employed in this thesis are presented in Chapter 5. Chapter 6 presents the discussions and comparisons of the performance results of our proposed technique along with regression technique and the standalone existing neural network techniques. Conclusion, summary of findings and future works are presented in Chapter 7.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter discusses the review and the leading preamble of linear and nonlinear systems. The appraisal focuses on regression with outliers in data sets, WLS regression methods, minimum covariance determinant estimator for location and dispersion in the presence of outliers as well as robust estimators for multiple regression analysis. The review will also hinge on the $CFBNN$ methods.

A system is a set of detailed methods, procedures and routines created to carry out a specific activity, perform a duty or solve a problem. In this study we will referred to a system as logical combination of a set of variable with the aim of performing certain functions that changes over time. Dissimilarity between a system and a model is a very vibrant idea in system identification. Scientifically, depiction of a system is referred to as a model. Like day to day life situation, a system may not necessary be pronounced perfect and will consequently comprise of errors. However, any related insinuation to the system persistently imitates the actual true system, hence, we can refer to model errors instead of system errors. A linear system is a system that constitute a constant or product of a parameter and a predictor variables with one basic form such that it can be expressed as a linear combination. A nonlinear system is a system that is unlike the linear system, takes many different forms such that it cannot be expressed totally as a linear combination.

2.2 Linear Methods

This study as earlier mention in Section 1.4 is focus on modeling and predicting/estimating performance of linear and nonlinear models. The linear models to be reviewed here are regression with outliers, WLS regression model, minimum covariance determinate estimator for location and dispersion in the presence of outliers and robust estimators for multiple regression analysis.

2.2.1 Regression with outliers

The impact of linear regression in the presence of outliers in data have been widely deliberated. Using linear regression methods in predicting the performance of a model with outliers data set can fail dramatically (Carroll and Ruppert, 1988). The dramatically failure of the performance of these predictions is as result of low variability regions having significant less influence setting parameters in making predictions than regions containing high variability (Payne, 2014). Linear regression in a outlier data thwarts (prevents) the existence of Type 1 errors as well as the probability coverage of confidence intervals (CI) for predictions of model-based from attaining the nominal level that result to the declaration of a statistical outcome significant in which it is not (Lim et al., 2010; Víšek, 2011). The widely used methods in dealing with outliers in data set are the transformations of the response variable and weighting. Transformation of correct response of variable results effectively to homoscedastic relationship (Fleiss, n.d.; Rasmussen, 1989; Luh, 1993). Gelfand (2015) in his study on the predictive ability of modern regression methods in understanding the impact of heteroscedasticity in data sets introduced the transformation of the response variable to correct heteroscedasticity with the focus on variance that increases as a result of the power of the mean. The log transformation was implemented in his study and he compared the predictive accuracy with the aid of evaluating metric of MSE. The comparison focuses on the log transformation with and without a variance stabilizer. Nine (9) regression methods was employed in simulating the predictive ability of his study. In order to understand the prevalence and magnitude of his methods, 42 real data sets were used to test each of the nine regression methods used in his study. The nine methods used in his study are linear regression, stepwise linear regression, least absolute selection and shrinkage operator, regression trees, random forests, boosted regression trees, multivariate adaptive regression splines, *ANN* as well as the Bayesian additive regression trees. These methods were applied in a generated linear and non-linear outlier data sets. Findings from his study revealed that despite the assumption of linearity in homoscedasticity, the linear methods (linear regression, stepwise linear regression and the least absolute selection and shrinkage operator) outperformed the modern methods (regression trees, random trees, random forests, boosted regression trees, multivariate adaptive regression splines, *ANN* and the Bayesian additive regression) in linear outlier data. However, the *ANN* method came out to be the best in almost all the cases considered regardless of the form of the error variance or data linearity.

Detections of bilinearity, correlation and heteroscedasticity are very important in regression analysis (Yingan and Bocheng, 2008). Whenever data are been collected in statistical analysis, the usual assumption is that all the random errors are expected to be mutually independent and should have an equal variances. This assumption fails in situations where data are collected sequentially over time which in turn raise to substantial serial correlation in the errors (Seber and Wild, 1989). Lindgren (2010) employed a two way alternative approach to lessen the problem of heteroscedasticity in dyadic regression models. The two way alternative considered by him involved the quadratic assignment procedure (QAP) and two-way clustering technique (TWC). In order to appraise the performance of his developed techniques, an extensive Monte Carlo experiment was adopted as demonstrative instances. He found that the consequences of not correcting the presence of outliers in data sets especially in a distorted Type 1 errors can be quite substantial whenever the error variances vary across the dyadic regression. Trial outcomes revealed that the developed alternative approaches performs better using regression techniques to study dyadic data and are well advised to correct heteroscedasticity. Outliers in small sample data have low power (Long and Ervin, 2000) as a result makes the detection of outliers in such sample data difficult.

Outliers in linear regression models can be considered as a problem in a context of group experiments which serve as a fertilizer trials where the error variances are ordered (Hoferkamp and Peddada, 2002). Lim et al. (2012) are of the view that uncertainty estimates of estimated parameters rest on the underlying structure of the error variance in a model. According to Lim et al. (2012), outliers and influential observation in toxicological data sets are often very common, and parameters of a nonlinear regression models are often used by toxicologists and pharmacologists in describing the toxicity of a chemical. Therefore, estimation of parameters of a nonlinear regression model becomes an important problem. They developed a robust estimation procedure called the preliminary test estimation (PTE). The PTE used two forms of M-estimation, the ordinary M-estimation (OME) and the weighted M-estimation (WME). Findings from the performance of their proposed estimator when compared with weighted M-estimation (WME) and ordinary M-estimation (OME) using simulation data revealed that, in outlier data, the PTE of WME gives a better result compared to the PTE of

OME. Zhu et al. (2016) in their study of reduction of dimensional based-regression models for testing heteroscedasticity, asserted that, for any analysis to take place, there is the need to check for the presence of outliers in the data to be use for the analysis. Regression analysis that ignored the checking of outliers may result in inaccurate inferences, such as, inefficient or even inconsistent estimates. Their proposed dimension reduction-based model adaptive test $(DRMAT)$ used the basic idea of constructing a test based on local smoothing test in a situation where a model-adaptive technique is utilized as proposed by Zheng (2009). Zhu et al. (2016) further asserted that their test construction did not only used the model structure of conditional variance, but also used the dimension reduction structure of the mean function. Their approach consider a general regression model of the form:

$Y = g(B_1^T X) + \delta(B_2^T X)e$, where $\delta(B_2^T X)e = \varepsilon$, B_1 is $p \times q_1$ matrix with q_1 representing the orthonormal columns, B_2 is $p \times q_2$ matrix, with q_2 represnting the orthonormal columns, e is independent of X, with $E(e/X) = 0$ and the function g and δ are unknown.

Under the model equation considered by Zhu et al. (2016), a null hypothesis H_0 and alternative hypothesis H_1 were constructed as follows:

$$H_0 : P\{Var(\varepsilon/X) = Var(\varepsilon/B_1^T X) = \sigma^2 = 1\} \tag{2.1}$$

for some σ^2, and

$$H_1 : P\{Var(\varepsilon/X) = Var(\varepsilon/B_1^T X) = \sigma^2 < 1\} \tag{2.2}$$

for all σ^2.

They further elucidated the advantage gain in their method to include *DRMAT*, irrespective of heavy computational burden. They computes critical values applying its limiting null distribution that is often an inherent property of local smoothing testing methodologies. More importantly to their embedded dimension reduction is the used of model-adaptive which allowed the use of more information on data in order to make the test so omnibus. The *DRMAT* has a very significant faster convergence rate of $O(n^{\frac{1}{2}}h^{\frac{q_1}{4}})$ to its limit than $O(n^{\frac{1}{2}}h^{\frac{p}{4}})$ under the H_0 as defined in equation (2.1) when $q_1 \leq p$ in existing tests, where h stands for the measure of the complexity. Their method also detect the local H_1 converging to the hypothesis at a very much faster rate of $O(n^{-\frac{1}{2}}h^{-\frac{q_1}{4}})$ than the typical rate of $O(n^{-\frac{1}{2}}h^{-\frac{p}{4}})$.

2.2.2 Weighted least squared (WLS) regression models

A *WLS* as defined by Guo et al. (2014) is a process in estimation techniques where the observations are weighted proportionally to the reciprocal of the error variance for that observation and in turn deal with the issue of non-constant variance. Consider the multiple linear regression model given by:

$$Y_i = \beta_o + \beta_1 X_{1i} + ... + \beta_j X_{ji} + ... + \beta_p X_{pi} + \varepsilon_i, \forall i = 1,2,...,n; j = 0,1,2,...,p+1 \quad (2.3)$$

where Y_i is the dependent variable, X_{ji} are independent variables, β_j are unknown parameters and ε_i is the error term.

Again, consider the following assumptions from a multiple linear regression (MLR):

(i) $E(\varepsilon_i) = 0$; for all i