AN IN-DEPTH STUDY OF TYPICAL MACHINE LEARNING METHODS VIA COMPUTATIONAL TECHNIQUES

Patrick Ozoh [1], Moruf Olayiwola [2], Adepeju Adigun [1]

[1] **Osun State University - Nigeria, Department of Information and CommunicationTechnology**
[2] **Osun State University – Nigeria, Department of Mathematical Sciences**

Corresponding Author: Patrick Ozoh, patrick.ozoh@uniosun.edu.ng

ABSTRACT: The ability to model and perform decision modeling and analysis is an essential feature of many real-world applications ranging from emergency medical treatment in intensive care units to military command and control systems. Models are essential in providing support for businesses processes, systems and dealing with complex problems. The development of appropriate models for planning and management is a tool for improving efficiency in real world problems. Consequently, a thorough study of common machine learning techniques is undertaken with the aim of comparing the techniques and identifying a suitable technique that is applicable to modeling and forecasting real world data. Modeling helps make informed decisions, using techniques for analysis, estimation and, forecasting. There is a lot of research published about machine learning techniques, with the intention of developing models for estimations. In order to identify an appropriate machine learning technique, it is necessary to carry out a comparative study of commonly used machine learning techniques. For this purpose, a review of Box-Jenkins technique, regression method and artificial neural network (ANN) is undertaken with the aim of identifying a reliable and accurate technique for modeling data.
KEYWORDS: machine learning, models, planning, estimation, efficiency.

1. INTRODUCTION

The analysis and modeling of real world problems are important features for performing decision making in diverse applications. This ranges from manufacturing applications, construction, military applications, health applications, logistic, transportation distribution, to mention just a few. In order to apply appropriate machine learning tools to modeling problems, it is expedient to analyze various techniques to identify the method that will produce accurate estimates ([R+10]). Nogales et al., ([N+02]) indicated that the introduction of machine learning technique in the analysis of problems would minimize modeling errors, since this has become a tool for model estimation, thereby improving efficiency in the modeling process.

A study to develop models was carried out using a combination of evolutionary local kernel regression (ELKR) model and Box Jenkins approach utilizing auto regressive integrated moving average method ([KS10]). The technique was used to analyze load forecast behavior on real power consumption for large-scale scenarios in smart grids. An analysis may be obtained by considering individual series as components of a multivariate or vector time series and analyzing the series jointly. This involves the development of statistical models and methods of analysis that adequately describe the inter relationships among the series. The model generated by a time series model estimated synthetic activity sequences based on patterns generated by the systems. This makes it necessary for decision making to be inferred from models in order to improve efficiency ([Y+00]).

2. DESCRIPTION OF TYPICAL MACHINE LEARNING TECHNIQUES

An in-depth study of machine learning techniques for modeling is undertaken, with the aim of identifying a reliable technique for modeling and forecasting data. The techniques presented in this study include Box-Jenkins, regression analysis and artificial neural network (ANN) techniques. Guidelines are needed for choosing the most appropriate technique to provide accurate estimates using these techniques. These techniques are discussed in details in the following sections.

2.1 Box-Jenkins Technique

The Box-Jenkins time series technique is presented in this section. The Box-Jenkins technique used in this study applies autoregressive integrated-moving average (ARIMA) process as a time series method. The idea of using a mathematical model to describe the behavior of a physical phenomenon is well established. In particular, it is possible to derive a model based on physical laws, which it is possible to calculate the value of some time-dependent quantity nearly exactly at any instant of time. If exact calculation were possible, such a model would be

entirely deterministic. Nevertheless, it may be possible to derive a probability or stochastic model, which models the probability of the future behavior of a value lying between two specified limits. The models for time series are stochastic models. A time series $z_1, z_2, ..., z_n$ of N successive observations is regarded as a sample realization from an infinite population of such time series that could have been generated by the stochastic process. The *backward shift operator* E is applied to the computation of Box-Jenkins technique and is defined by $Bz_t = z_{t-1}$.
Hence, $B^n z_t = z_{t-n}$
Also,

$$\nabla z_t = z_t - z_{t-1} = (1-B)z_t$$

The stochastic models employed are based on the idea that an observable time series z_t is transformed to the process a_t, with $\psi_1, \psi_2, ...$ as weights. i.e.

$$z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + ... = \mu + \psi(B)a_t \quad (1)$$

In general, $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + ...$
Let $t, t-1, t-2, ...$ be $z_{t-1}, z_{t-2}, ...$
Also let $\tilde{z}_t = z_t - \mu$ be the series of deviations from μ The autoregressive model given by Equation (1) may be written economically as $\phi(B)\tilde{z}_t = a_t$
The model contains $p+2$ unknown parameters $\mu, \phi_1, \phi_2, ..., \phi_p, \sigma_a^2$.
The autoregressive (AR) model is a special case of the linear filter model of Equation (1), for example, $\tilde{z}_{t-1}$ can be eliminated by substituting

$$\tilde{z}_{t-1} = \phi_1 \tilde{z}_{t-2} + \phi_2 \tilde{z}_{t-3} + ... + \phi_p \tilde{z}_{t-p-1} + a_{t-1}$$

Similarly, $\tilde{z}_{t-2}$ can be substituted. Hence,

$$\tilde{z}_t = \phi^{m+1}\tilde{z}_{t-m-1} + a_t + \phi a_{t-1} + \phi^2 a_{t-2} + ... + \phi^m a_{t-m}$$

Symbolically, in the general AR case, we have that $\phi(B)\tilde{z}_t = a_t$ is equivalent to

$$\tilde{z}_t = \phi^{-1}(B)a_t = \psi(B)a_t$$

with $\psi(B) = \phi^{-1}(B) = \sum_{j=0}^{\infty} \psi_j B^j$

The general form of the model that is used to describe time series is the ARIMA model

$$\varphi(B)z_t = \phi(B)\nabla^d z_t = \theta_o + \theta(B)a_t \quad (2)$$

where
$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_P B^p$
$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$
$\phi(B)$ and $\theta(B)$ are polynomial operators in B of degrees p and q. This process is referred to as an ARMA (p, q) process. The ARIMA model can be expressed explicitly in terms of current and previous shocks. A linear model can written as the output z_t from the linear filter

$$z_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + ... = a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j}$$

$$= \psi(B)a_t \quad (3)$$

whose input is a white noise, or a sequence of uncorrelated shocks a_t with mean 0 and common variance σ_a^2. Operating on both sides of Equation (3) with the generalized autoregressive operator $\varphi(B)$, then,

$$\varphi(B)z_t = \varphi(B)\psi(B)a_t,$$

However, since $\varphi(B)z_t = \theta(B)a_t$, it follows that

$$\varphi(B)\psi(B) = \theta(B)$$

Therefore, the ψ weights can be determined in the expansion

$$(1 - \varphi_1 B - ... - \varphi_{p+d} B^{p+d})(1 + \psi_1 B + \psi_2 B^2 + ...)$$
$$(1 - \theta_1 B - ... - \theta_q B^q) \quad (4)$$

Thus, the ψ_j weights of the ARIMA process can be determined recursively through the equations

$$\psi_j = \varphi_1 \psi_{j-1} + \varphi_2 \psi_{j-2} + ... + \varphi_{p+d} \psi_{j-p-d} - \theta_j, j > 0$$

with $\psi_0 = 1, \psi_j = 0$ for $j < 0$, and $\theta_j = 0$ for $j > p$. It is noted that for j greater than the larger of $p+d-1$ and q, the ψ weights satisfy the homogenous difference equation defined by the generalized autoregressive operator, that is,

$$\varphi(B)\psi_j = \phi(B)(1-B)^d \psi_j = 0$$

where *B* now operates on the subscript *j*. Thus, for sufficiently large *j*, the weights ψ_j are represented by a mixture of polynomials, damped exponential and damped sinusoid in the argument *j*.

2.2 Regression Analysis

Regression techniques belong to the class of causal models. The regression technique, as described by Wu (2018), is discussed in this section. Regression is the study of relationships among variables, a principal purpose of which is to predict, or estimate the value of one variable from known or assumed values of other variables related to it. To make predictions or estimates, we must identify the effective predictors of the variable of interest. A regression using only one predictor is called a simple regression. Where there are two or more predictors, multiple regressions analysis is employed. To develop a regression model, begin with a hypothesis about how several variables might be related to another variable and the form of the relationship, i.e.
Given a line $y = mx + b + \varepsilon$, and letting

$$\varepsilon^2 = \Sigma (y_i - y_p)^2 \qquad (5)$$

y_p = predicted values, and y_i = actual values. Also, let

$$y_p = m\, x_i + b \qquad (6)$$

Combine Equation (5) and Equation (6),

$$\varepsilon^2 = \Sigma (y_i - m\, x_i - b)^2 = \Sigma (m^2 x_i^2 + 2mbx_i - 2mx_i y_i + b^2 - 2by_i + y_i^2) \qquad (7)$$

Equation (7) is at a minimum, i.e.

$$\frac{ds^z}{dm} = 0 \text{ , and } \frac{ds^z}{db} = 0 \qquad (8)$$

Solving for Equation (8), gives

$$\frac{ds^z}{dm} = 2m\Sigma x_i^2 + 2b\Sigma x_i - 2\Sigma (x_i y_i) = 0 \qquad (9)$$

Solving for Equation (8), gives

$$\frac{ds^z}{db} = 2m\Sigma x_i + 2\Sigma b - 2\Sigma y_i = 0 \qquad (10)$$

To simplify the notations, let

$Sx = \Sigma\, x_i$
$Sy = \Sigma\, y_i$
$Sxy = \Sigma\, (x_i y_i)$
$Sxx = \Sigma\, (x_i^2)$
$\Sigma\, b = nb$

From Equation (9),

$$Sxy = mSxx + bSx \qquad (11)$$

From Equation (10),

$$Sy = mSx + nb \qquad (12)$$

The optimized values for *m* and *b* are respectively given as:

$$m = \frac{n\sum x_i y_i - \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{\sum y_i}{n} - \frac{m\sum x_i}{n}$$

2.3 Artificial Neural Network (ANN)

The structure of an artificial neural network (ANN) is illustrated in Figure 1, while the model for an artificial neuron is shown in Figure 2. The ANN process is divided into: training, validation and testing. The definitions of these terminologies were discussed by Priddy & Keller (2005) and are given as follows:
Training set: A set of data used to modify weights.
Validation set: It consists of using errors to adjust the training set.
Testing set: This consists of using real data for analysis.

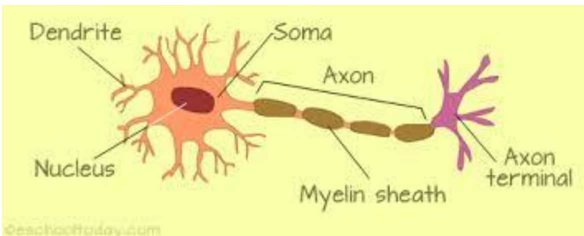

Fig. 1. Structure of a neuron

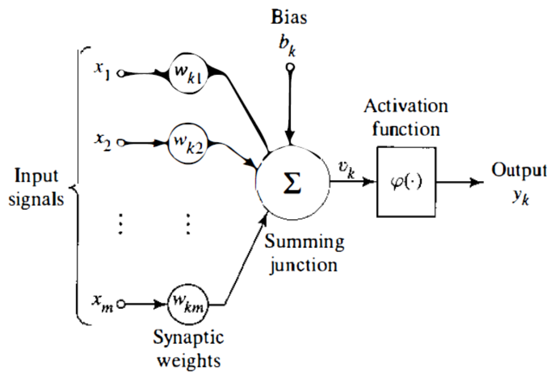

Fig. 2. Artificial neuron model

The model illustrated in Figure 2 is given in Equation (13) and Equation (14).

$$u_k = \sum_{j=1}^{m} w_j x_j \qquad (13)$$

and

$$y_k = \varphi\left(u_k + b_k\right) \qquad (14)$$

where $x_1, x_2, \ldots, x_m$ are the input signals; $w_1, w_2, \ldots, w_m$ are weights; u_k is output; b_k is the bias; $\varphi(.)$ is the activation function and y_k is signal for ANN**.** Also,

$$v_k = u_k + b_k \qquad (15)$$

where v_k is activated by k. The linear combiner u_k is modified by b_k (Figure 3).

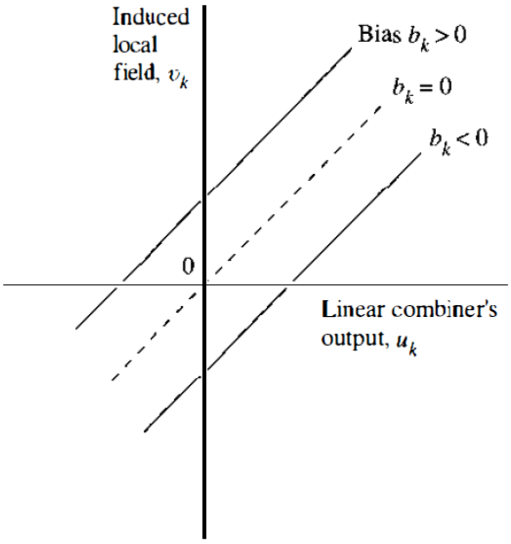

Fig. 3. ANN output

Artificial neural networks can be classified into:
(i) Feedforward ANN
(ii) Feedback/Recurrent ANN

i. Feedforward ANN

The topology of the Feedforward ANN is given by Figure 4.

ii. Feedback/Recurrent ANN

The Feedback/Recurrent ANN is based on results of previous processing inputs.

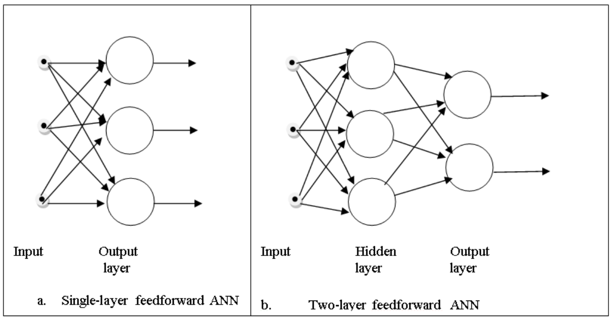

Fig. 4. Feedforward ANNs Topology

3. COMPARATIVE ANALYSIS OF THE TECHNIQUES

The inherent features of applying the machine learning techniques are summarized in Table 1. The techniques discussed in this section were examined to determine their strengths and weaknesses. The techniques summarized in this table include Box-Jenkins technique, regression analysis, and the artificial neural network (ANN). As a result, challenges in applying these techniques discussed and analyzed.

Table 1. Characteristics of machine learning techniques

Tech-nique	Positive Attributes	Negative Attributes
Box-Jenkins	1) Adapts multiple variables 2) Solution based on multi-variables 3) Simple input data. 4) Possible to adjust model to more variables 5) Discusses trends	1) Does not differentiate between components of variables 2) No explicit description of components of variables 3) Need to identify a more appropriate technique for modeling 4) Does not specify requirements for individual variables 5) No information of specific effect of individual variables on dependent variable
Regression Analysis	1) Incorporates multiple variables 2) Information on effect of individual variables on dependent variable 3) General mathematical model to estimate variables 4) Handles larger data than Box-Jenkins technique 5) Model shows when variables are included or not	1) Small and large transitions in time are under-represented in the model 2) No explicit description of components of individual variables 3) Does not differentiate between components of variables

Artificial Neural Networks	1) Discusses the use of multiple variables 2) Introduces artificial intelligence for model 3) Discusses modelling by validating and testing data. 4) Determines model's behaviour. (5) Handles large data	1) Does not specify model requirements for individual variables 2) No information of effect of individual variables on dependent variable.

4. CONCLUSIONS

An in-depth study of commonly used machine learning techniques has been carried out. Each technique has its own limitations and challenges. Modeling data requires a high degree of accuracy such that any deviation from expectation may result in obtaining accurate estimates. The techniques used in the various studies were existing machine learning techniques applied to heterogeneous problems. Based on the reviewed papers, the following challenges were identified for Box-Jenkins, regression analysis, and artificial neural networks (ANN):
1. The need to identify a more accurate and generic technique for predicting data.
2. The need for models to handle big data.
3. Lack of insight into the estimation of components for variables.
4. There is a need to apply a technique for modeling.
5. The study of the extent to which various components affect variables.
The artificial neural network (ANN) has being able to overcome all the challenges present for Box-Jenkins technique and regression analysis. The ANN technique is more reliable for predicting data compared with Box-Jenkins technique and regression analysis. The ANN technique will also help identify the contribution of components for the various variables that are considered in a study. The application of machine learning techniques to model data are related to model behavior and their interventions in various fields of study, including information technology. This research reveals similarities and differences, such as the characteristics of machine learning techniques. These reviewed studies have added to our understanding of how to model data using the techniques mentioned in this study.

REFERENCES

[Kal00] **S. Kalogirou** - *Applications of artificial neural-networks for energy systems*, Applied energy, vol. 67: 17-35, 2000.

[KS10] **B. Kramer, J. Satzger** - *Power prediction in smart grids with evolutionary local kernel regression*, Hybrid artificial intelligent systems, vol. 1(1): 262-269, 2010.

[N+02] **F. J. Nogales, J. Contreras, A. J. Conejo, E. Espinola** - *Forecasting next-day electricity prices by time series models*. IEEE Transactions on Power Systems, vol. 17(1): 342–348, 2002.

[Pat96] **D. Patterson** - *Artificial neural networks*. Singapore: Prentice Hall, 1996.

[PK05] **K. Priddy, P. Keller** - *Artificial neural network - An introduction*, USA SPIE international society for optical engineering, 2005.

[R+10] **A. Ruzzelli, C. Nicolas, C. Schoofs, G. O'Hare** - *Real-time recognition and profiling of appliances through a single electricity sensor*, 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communication and Networks, 2010.

[Wu18] **C. Wu** - *Regression Technique*, Retrieved from http://www.historyofinformation.com/expanded.php?id=2706, 2018.

[Y+00] **S. Yao, Y. Song, L. Zhang, X. Cheng** - *Wavelet transform and neural networks for short-term electrical load forecasting*. Energy conversion and management, vol. 41(18): 1975–1988, 2000.