Probability-based Non-Linear Model: Approach to Population Forecasts for Proper Development Planning in Nigeria

Ismaila W. Oladimeji
Ladoke Akintola University of Technology, Ogbomoso
E-mail: woismaila@lautech.edu.ng

Akinola, Lukman S.
Fountain University, Osogbo, Nigeria. E-mail: lukmanakinola@gmail.com

Akinnuwesi Boluwaji. A.
Bells University of Technology, Ota, Nigeria. E-mail: moboluwaji@gmail.com

Falohun Adeleye. S.
Ladoke Akintola University of Technology, Ogbomoso, Nigeria.
E-mail: asfalohun@lautech.edu.ng.

ABSTRACT

Despite the series of population census that has been carried out over the years, the question of how to forecast relatively accurate population for the subsequent years for proper planning in Nigeria still remains under investigation. Non-linear model (Neural Network; NN) has been developed and predictions are carried out on past population data. The results were compared with conventional method (Time series, TS) and it is found that the NN model performs better than the Time series model.

Keywords: *Nigeria, Multi-layer feed-forward, Back-propagation, Time Series, Population*

INTRODUCTION

The population of a country is key factor in the developmental process of any nation, considering it in terms of quality and quantity. Population is the total number of people living in a country or particular geographical area at a given time. Population can also be described as the analytical study of the sex, age, and geographical spread of people living in a particular country. The importance of population management in any nation whether developed or developing is so obvious. Today, United Nation alerts all the nations of the world on the need to manage its population primarily to aid development in all areas of human habitation on the earth (Ruggles, 1992). This then calls for the need to have a reliable and efficient archive of information and data. The National Population Commission was then created by the Federal Government of Nigeria and charged with huge responsibility of providing information on issues that centers on population management. Such issues include morbidity, mortality, fertility rate, growth rate, level of literacy, etc. amongst others (www.population.gov.ng/index.php/population-programmes/population-policy). National development cannot be achieved except the government and stakeholders are well informed about trends in population issues as mentioned

above. In view of this, information and data processing is a very important phenomenon that needs urgent attention. This is to assist government and policy makers in making decisions that will assist the nation in achieving its developmental goals and objectives. As such, this seeks the interest of the researcher in demonstrating the population growth trends in Nigeria using data mining technology. Population forecasting is now part and parcel of public administration and private business. The advent of electronic technology in data manipulation and the blossoming of statistical modeling have no doubt been factors in its growth. Forecasting the trends in human population is a complex problem.

Traditionally, statistical approaches are used to forecast the demography of a country. But, they are not very suitable for predicting a chaotic system like population. Firstly, statistical approaches make some assumptions, which are sometimes found unrealistic (Keilman *et al,* 2002). Secondly, statistical population forecast procedures cannot deal with the intrinsic chaos. But, to analyze the population data and to predict their trend in a given country requires the application of some technique that can deal with the intrinsic chaotic nature. Introduction of Artificial Neural Network (ANN) has immensely empowered the forecasting techniques of intricate systems such as Electrical load forecasting (Chaturvedi *et al,* 2003); meteorological forecasting (Smith et al 2005) etc. But, a literature survey shows that despite immense intricacy, population has not been the area of application of ANN. But Yuan and Hsu (2006) implemented ANN in analyzing the impact of population upon industrial structure.

This work employed time series and neural network based on feed-forward Multilyer Perceptron and Backpropagation models to predict the population of Nigeria. Each year, Nigeria adds more people to the world's population than many other countries. Nigeria population trend is complex attributable to various reasons. The reasons can be summarized as follows: Fertility rate has declined but the number of women in their reproductive age has increased rapidly; States of Nigeria vary significantly with respect to fertility, mortality and contraceptive use; Since independence, average life expectancy has increased substantially; Infant mortality rate has decreased over the years since independence; Nigeria is facing increased rate of HIV and other sexually transmitted disease cases like other third world countries.

All the features described above add different degrees of complexity to the population structure of Nigeria and thereby make the prediction task more convoluted. Conventional statistical procedures cannot take care of the chaos contributed by the aforesaid factors. Thus, some methodology is essential that can deal with the non-linearity and chaos intrinsic in Nigeria population. To the best of the knowledge of the authors of this work no significant work is available on predicting Nigeria population through modern mathematical methodology like neural network, genetic algorithm, etc.

MATERIALS AND METHOD

Time series: Time series is an ordered sequence of values of a variable at equally spaced time intervals. The factors are not affected for this forecasting but the predicting depended on the past valuable such as the growth in sale, gross national product or stock market analysis. The example of time series methods are Autoregressive Moving Average (ARMA), Exponential Smoothing, Extrapolation, Linear Prediction, Trend Estimation, Growth Curve and Box-Jenkins. Time series is employed in this work (Thompson *et al*, 1989)

Cause-And- Effect Model: Cause-and-effect model assumes that the variable to be forecasted is exhibited by the explanatory relationship with one or more independent variable. The purpose of this model is to discover the form of the relationship and to forecast future of the variable. A simple regression model contains only one independent (explanatory) variable, X_i, for i = 1,...., n subjects, and is linear with respect to both the regression parameters and the dependent variable. Regression Analysis can be applied for several purposes including data description, parameter estimation and control. Linear regression applications are adopted almost every fields, engineering, physical sciences, economics management and so on (Baker and Tahir, 2009).

The other technique of cause-and-effect model is neural network. This method is used for many applications such as financial, exchange rate, water flow, and inventory demanding and so on. The main characteristic of neural network is the ability to learn from environment, and to improve its performance through learning (http://leenissen.dk/fann/report/node4.html). Table 1 below shows the superiority of neural network on other methods.

Table 1: Comparisons of Past works on forecasting Models

Authors	Methodology	Contributions to research
Zhang, 2001	Neural Network, ARIMA	-Neural network outperforms than linear time-series model. -Input and hidden nodes can affect neural network modeling.
Pulido-Calvo, Montesinos, Roldan and Ruiz-Navarro, 2007	Linear multiple regression and feedforward computational neural network (CNNs)	-CNNs performed better than the regressions when water demand and climatic variables were considered as input data.
Baker and Tahir, 2009	Multiple linear regression, MLP	-MLP is more accuracy than regression but more complex, while regression is quite simple.
.Noori, Khakpour, Omidvar and Farokhnia, 2010	Multiple linear regression, MLP	-MLP model is the satisfactory predicting performance.

Mean Absolute Percent Error (MAPE)

The performance of each model is evaluated by using Mean Absolute Percent Error (MAPE) criteria. MAPE can be expressed as the following equations:

$$\text{MAPE} = \frac{1}{N}\sum_{t=1}^{n}\frac{|P_t - D_t|}{D_t} \qquad \ldots\ldots\ldots\ldots\ldots\ldots 1$$

where, P_t and D_t are the predicted and observed values of output respectively; N is the number of observations or time periods over which the errors are computed. A model with the minimum error is considered the best choice for prediction. (Lee and Miller, 2001)

A neural network (NN) model is composed of many artificial nodes that are linked together. The objective is to transform the inputs into meaningful outputs. NNs have the ability to learn by example.

Capabilities of ANNs include: pattern classification, clustering, function approximation, forecasting and optimization. An ANN is formed by a large number of processing neurons interconnected by weights. ANNs are classified into feed forward and recurrent networks. In a feed forward network, neurons are grouped into layers and the signals flow from one layer to another in the forward direction. Multi-Layer Perceptron (MLP) and radial basis function (RBF) networks are feed forward networks. In this study, the most commonly used NN model- the four-layer feed-forward NN model with Back-propagation method (BPNN) - is used for prediction.

A typical feed forward network as shown in figure 1 consists of an input layer, a hidden neuron layer and an output layer of neurons. Input layer simply transmits inputs through weighted links to hidden neurons where weighted inputs are accumulated and processed by a transfer function to generate an output to be sent to the output layer. A similar process takes place in the neurons in the output layer where outputs are generated.

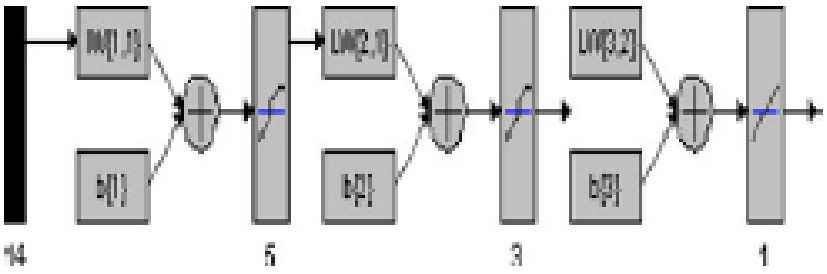

Figure 1: Topology of 4-layer MLFF Network.

The network learns the problems by taking partial derivatives of the weights and then applying them to each of the weights which takes place starting from the output layer to hidden layer weights, then the hidden layer to input layer weights, (this is necessary since changing these set of weights requires that the partial derivatives calculated in the layer downstream are known) this algorithm has been called the "back propagation algorithm". In the BPNN architecture, each node at input and hidden layers receives input

values, processes and passes to the next layer. This process is conducted by weight which is the connection strength between two nodes. The number of nodes in the input layer and the output layer are determined by the number of input and output parameters. (http://leenissen.dk/fann/report/node4.html). The back-propagation method algorithm is as followed:

i. Present the input vector patterns to the network;

ii. Propagate the signals forwards, and calculate

$$u_j = a_{0j} + \sum_{i=1}^{I} a_{ij} x_i, \quad v_k = b_{0k} + \sum_{j=1}^{J} b_{jk} y_j,$$
$$y_j = g(u_j), \quad j = 1,\ldots,J, \quad z_k = g(v_k), \quad k = 1,\ldots,K \qquad \ldots\ldots\ldots 2$$

iii. Calculate the mean squared error

$$E = \frac{\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}(z_{kn} - t_{kn})^2}{NK} \qquad \ldots\ldots\ldots\ldots 3$$

iv. Update the weights according to the delta rule:

$$w^{m+1} = wm - \lambda d^m, \quad d^m = \sum_{n=1}^{N}\left(\frac{\delta y}{\delta x}|_m\right)_n \qquad \ldots\ldots\ldots\ldots\ldots 4$$

v. Repeat the above steps ii, iii, iv until the error is less than the predefined value or for a predefined number of iterations.

After configuration input parameters of neural network models, next step is to train neural network models with these settings. In present case, the models are trained by using the following train data format: $x_i = f(x_{t-1},..,x_{t-n})$ where; x = population number of a particular year, n = number of input nodes in the model, t = desired year, f(x) is training function of neural network models.

Generally, the number of nodes in the input layer depends on the number of possible inputs used in the model, while the number of nodes in the output layer depends on the number of desired outputs. The number of hidden layers and how many nodes in each hidden layer cannot be well defined in advance.

Error Estimate

For a given input vector, the output vector is compared to the correct result. If the difference is zero, no learning takes place; otherwise, the weights are adjusted to reduce this difference. The learning is done by least-square-error minimization. The least- square -error (E) between the target output (T) and actual output (O) is given by:

$$E = \frac{1}{2}(T_i - O_i)^2 = \frac{1}{2}[T_i - f(w_i x_i)]^2 \qquad \ldots\ldots\ldots 5$$

Where, w_i ⇒ The weight matrix associated with i^{th} neuron

x_i ⇒ Input of the i^{th} neuron

O_i ⇒ Actual output of the i^{th} neuron

T_i ⇒Target output of the i^{th} neuron

The activation function f(x) is taken as the sigmoid function

$$f(x) = \frac{1}{1 + e^{-x}} \quad \ldots\ldots\ldots\ldots 6$$

The error gradient vector is

$$\nabla E = -(T_i - O_i) f'(w_i x_i) x_i \quad \ldots\ldots\ldots\ldots 7$$

Since minimization of the error is the objective, the weight is updated as

$$\Delta w_i = -\mu \nabla E \quad \ldots\ldots 8$$

Where μ is a positive constant. Hence, becomes; Δw_i

$$\Delta w_i = \mu (T_i - O_i) f'(w_i x_i) x_i \quad \ldots\ldots 9$$

Applying discrete mathematics, the weight vector is updated as:

$$w_i(k+1) = w_i(k) + \mu (T_i - O_i) f'(w_i x_i) x_i \quad \ldots\ldots\ldots\ldots 10$$

To frame the input matrix for the NN, the population is grouped into two viz (i) grouping the populace into age groups and (ii) total populace for each year. The grouping of ages include 0 to 8, 7 to 14, 12 to 19, 16 to 34, 15 to 49, 20 to 57, 35 to 69, 47 to 79, 80+. Instead of dividing the whole population into disjoint subclasses as done in the original data, intersecting subclasses of age are considered to make the data set more suitable for predicting a chaotic system. Male and female population (in thousands) in each subclass, married female population (in thousands) in each subclass, average life expectancy of male, average life expectancy of women, and total fertility rate are considered for the years 1998 and 2010 from U.S. Bureau of the Census, International Data Base. (http://www.motherlandnigeria.com/population.html"). The purpose is to predict the population (in each age group) in the year 1998 and 2010 only. Also, the population census for year 1907 to 2050 is used for predicting subsequent years.

RESULTS AND DISCUSSION

This study uses Nigeria population data as described in section 3. Time series model is implemented in Statistical Product and Service Solutions (SPSS) package version 22 and neural network (NN) is simulated in MATLAB 7.0 environment. Table 2 below describes the performance statistics of neural network models by using test data sets. For choosing the optimal architecture of neural network models, 100 test cases are made. Forecast errors of 100 neural models are calculated in the commonly used performance measure; MAPE.

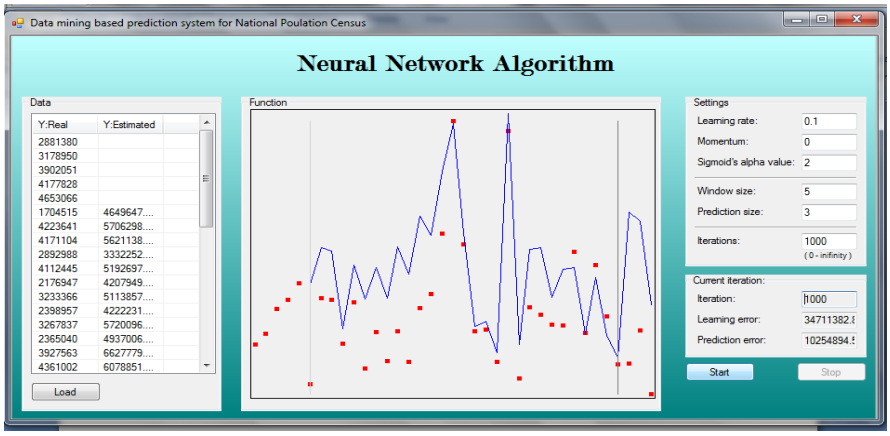

Fig 2. Neural Network screen after prediction

Table 2: Forecasting Performance of Time Series and ANN models for in-sample and out-of-sample period

Years	$MAPE_{TS}$	$MAPE_{ANN}$
1950-2006	1.238	0.0213
1990-2006	1.185	0.0075
1950-2015	2.033	0.0084
1990-2015	1.004	0.0153
1950-2050	2.072	0.0477
1990-2050	1.037	0.0748
1998(Age range)	2.718	0.1480
2010(Age range)	2.484	0.1250

In Table 2, we present the MAPE measures for the in-sample and the out-of-sample period respectively of the estimated models. There is no case that TS model outperforms the ANN model in both cases. It is however noticed that the prediction errors are more pronounced in the years that include 1950. This might be due to five ten years difference from 1950 to 1990 while 1990 to 2006 has just a year interval. The data predicted by U.S. Bureau of the Census, International Data Base (http://www.motherlandnigeria.com/population.html) from 2007 to 2050 was compared with the predicted values of both models. However, the MAPE results produced by ANN are close to zero for in-sample and out-of-sample period which means that the ANN produced a better forecast than TS.

CONCLUSION

Forecast experiments were conducted for population in Nigeria. Statistics tool, MAPE was calculated to examine the performance of these experiments. From experimental results, it is apparent that prediction errors are found to be very less than one in non-linear NN as a predictive model for population in Nigeria. Changes to number of nodes of neural network model do not affect the

performance of the model. The results reflect that the application of neural network models is feasible for population prediction model of Nigeria.

REFERENCES

BackPropagation Neural Network model; http://ieee.uow.edu.au/~daniel/software/libneural/BPN_tutorial/BPN_English/BPN_English/node8.html, 10/4/2012, 7:43PM.

Baker, N. M. and Tahir, I.M. (2009), Applying Multiple Linear Regression and Neural Network to Predict Bank Performance. International Business Research, Vol.2, No.4, pp. 176-183.

Bandyopadhyay G. and Chattopadhyay S. (2008). An Artificial Neural Net approach to forecast the population of India.

Chaturvedi, D. K., M. Mohan, R. K. Singh and P. K. Kalra, (2003). Improved generalized neuron model for short-term load forecasting, Soft Computing 8: 10-18.

Keilman, N., Pham, D. and Hetland, A. (2002). Why population forecasts should be probabilistic - illustrated by the case of Norway. Demographic Research 6, Article 15.

Lee, R. (1993). Modelling and forecasting the time series of US fertility: Age patterns, range and ultimate level.International Journal of Forecasting 9, 187-202

Lee, R. and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality, "in: Demography 38 (4), 537-549

Neural Network Model; http://leenissen.dk/fann/report/node4.html, 9/25/2012, 11:18AM.

Noori, R., Khakpour, A., Omidvar, B. and Farokhnia, A. (2010), Comparison of ANN and principle component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistics. Expert Systems with Applications, 37, 5856-5862.

Pulido-Calvo, I., Montessinos, P., Roldan, J. and Ruiz-Navarro, F. (2007), Linear regression and neural approaches to water demand forecasting in irrigation districts with telemetry systems. *Biosystems Engineering*, 97, 283-293.

Ruggles, S. (1992). Migration, marriage and mortality: correcting sources of bias in English family reconstitutions. *Population Studies*, 46: 507-522.

Smith, B. A., Ronald W. M., and Gerrit H. (2005). An Enhanced Artificial Neural Network for Air Temperature Prediction, Transactions on Engineering Computing and Technology 7(August): 7-12.

Thompson, P.A., Bell, W.R., Long, J.F. and Miller, R.B. (1989). Multivariate Time Series Projections of Parameterized Age-Specific Fertility Rates. *Journal of the American Statistical Association*, 84 (407), 689-699

Yuan, B. J. and J. Hsu, (2006). Application of population structure change on industry structure forecast. IAMOT 2006: 1-15.

Zhang, G. P. (2001). An Investigation of neural networks for linear time-series forecasting. Computer & Operation Research 28, pp. 1183-1202.