American Journal of Engineering Research (AJER)
e-ISSN : 2320-0847 p-ISSN : 2320-0936
Volume-03, Issue-11, pp-131-138
www.ajer.org
Research Paper Open Access

A Neural Network Approach to GSM Traffic Congestion Prediction

M. A. Raheem[1, a] and O. U. Okereke[2, b]
[1, 2]*Department of Computer and Communication Engineering*
Abubakar Tafawa Balewa University Bauchi
Nigeria
[a]*marufakin@yahoo.com*, [b]*ouokereke@yahoo.com*

ABSTRACT : *In this paper, we propose a GSM congestion prediction model based on multilayer perceptron neural networks (MLP-NNs) with sigmoid activation function and Levenberg-Marquardt Algorithms (LMA) using twelve month real traffic data. The trained network model was used to predict traffic congestion along a chosen route. Regression analysis between predicted traffic congestion volumes and corresponding actual traffic congestion volumes shows a correlation coefficient of 0.986. This result clearly shows the effectiveness of Artificial Neural Networks (ANN) in traffic congestion prediction and control.*

KEYWORDS: *GSM congestion, prediction, neural networks, activation function, LMA*

I. INTRODUCTION

As wireless networks mature and the rate of subscriber growth levels off, the focus on customer retention and satisfaction becomes increasingly important. This occurs at a time when the widespread use of mobile communications has heightened consumer demand for quality service anytime, anywhere. Today, network operators face the challenges of improving the quality of service while increasing capacity and rolling out new services. Operators are fast realizing that they are in a highly competitive environment where subscribers can make or break them. Dissatisfaction by subscribers gives rise to a high rate of subscriber churn and low revenue for the operator.Congestion is a problem all GSM service providers are facing and trying to solve. It is a situation that arises when the number of calls emanating or terminating from a particular network is more than the capacity that the network is able to cater for at a particular time. It causes call signals to queue on the transmission channel. Consequently, the rate of transfer of voice signals is reduced or quality of signals received become distorted or both. At worst, the calls will not connect at all. There are technical mechanisms along the transmission link that tend to create or worsen congestion. When a Mobile Station (MS) dials a number, the call is routed to the nearest Base Transceiver Station (BTS) through the Air (Um) Interface. This receives, amplifies, and reroutes the call to the Base Station Controller (BSC) through the ‘Abis’ Interface. The BSC controls and manages multiple BTSs and communicates directlywith the Mobile service Switching Center (MSC) with an interface called the “A” Interface . Communication between two MSCs occurs on the ‘E’ Interface. The destination MSC finally routes the calls to its destination after the identity, authentication and credit status of the subscriber have been verified. The call set-up scenario is as shown in figure 1.

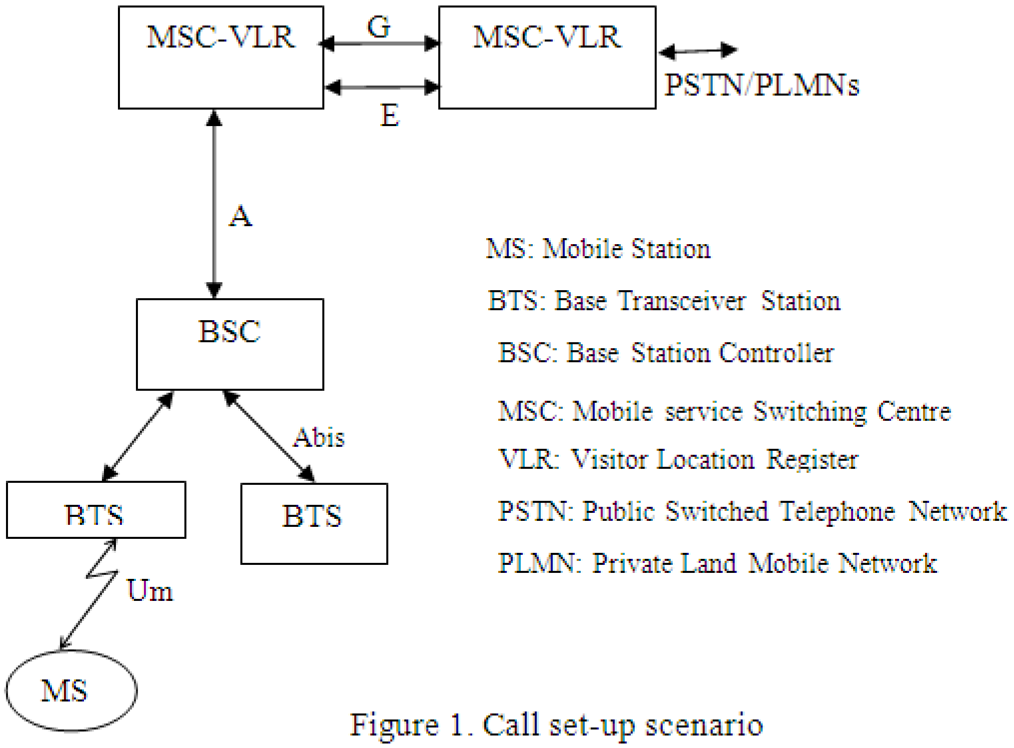

Figure 1. Call set-up scenario

Network Congestion : Congestion is defined as a state of network elements (e.g. switches, concentrators, cross-connects and transmission links) in which the network is not able to meet the negotiated network performance objectives for the already established connections and/or for the new connection requests (ITU-T, 1993). It is a network state in which performance degrades due to the saturation of network resources, such as communication links, processor cycles, and memory buffers (Andreas and Ahmet, 1999).From the view point of the network, four basic elements are related to congestion or indicate that a call could not be completed (Kuboye, 2006).

Traffic Channels Congestion (TCHC): Traffic channels (TCH) are voice channels. Out of the eight channels defined for each radio frequency carrier, most are used as traffic channels while some are used as control channels (Mehrotra, 1997). With absence of free voice channels during call initiation, traffic channels congestion (TCHC) results.

Dedicated Control Channel Congestion (DCCHC): Standalone dedicated control channels (SDCCH) provide authentication to mobile station, location updating and traffic channel assignments during idle periods (Mehrotra, 1997). Absence of free SDCCH for authentication results in an unsuccessful call attempt. This is called dedicated control channel congestion.

Common Control Channels Congestion (CCCHC): Common control channels support the establishment and maintenance of communication links between mobile stations and base stations (Harte et al., 1999). Random access channels (RACH), paging channels (PCH) and access grant channels (AGCH) are all common control channels. RACH is used for network assignment request, PCH for incoming call alert while AGCH assigns mobile stations to specific DCCH or SDCCH for onward communication. Calls cannot be established without the availability of any of these common control channels.

Pulse Code Modulation Congestion (PCMC Pulse code modulation (PCM) or E1 is the link required to connect the base station controller (BSC) and mobile service switching centre (MSC) or two MSCs together. Each PCM consists of thirty two timeslots. When there is no free timeslot to carry the call signals between the BSC and MSC or between the two MSCs (in case of inter-MSC call), then we have pulse code modulation congestion.

Congestion Control Strategies: Congestion control refers to the set of actions taken by the network to minimise the intensity, spread, and duration of congestion. It can be said that it is that aspect of a networking protocol that defines how the network deals with congestion. Despite the many years of research efforts, the problem of network congestion control remains a critical issue and a high priority, especially given the growing size and demand of the networks.Several attempts had been made to forestall and manage congestion in GSM/GPRS network. A dynamic channel allocation model with one-level buffering was formulated by Ojesami *et al.* (2011) to control congestion in GSM network with a view to prevent call loss or degradation in quality of service of calls using Markov chain technique. They used object oriented program to evaluate the performance of the scheme based on three performance metrics: resource utilization, average queue length and blocking probabilities. The results obtained from the model show that the proposed scheme provides better performance benefits over fixed threshold techniques.

Kuboye (2010) proposed some optimization techniques that could be applied to minimize the problem of congestion on the GSM network in Nigeria. It was argued that implementation of dynamic half rate decoding, national roaming agreement between operators, regionalization and merging of GSM networks could ameliorate the problem of congestion. Dynamic load balancing technique was combined with Call Admission Control (CAC) to re-route calls that would have been dropped to another less busy cell within the BSC area (Alarape *et al.*, 2011). The combined algorithms were implemented on JAVA platform using real life call data record (CDR) collected from Globacom Nigeria Limited. New Call Blocking and Handoff Call Dropping Probabilities (NCBP and HCDP) were employed as performance index. NCBP and HCDP were computed for both CAC only and the combined scheme. The obtained results show significant reduction in the values of both NCBP and HCDP by 71.43% and 100% respectively, of cells considered for the new combined scheme when compared with that of the CAC only. Thus, the new combined scheme enabled the cells to accommodate more calls thereby increasing the call carrying capacity of the network. Markus *et al.* (2011) used multilayered feed-forward neural network with gradient descent backpropagation algorithm to model the telephone traffic situation on a PSTN. Regression analysis between predicted traffic congestion volumes and corresponding actual volumes clearly shows the utility and effectiveness of neural networks in traffic prediction and congestion control.

Realistic traffic models in the mobile network should cater for both the user communication rate and user movement in the network (Lam *et al.*, 1997). Little work has been done to develop adequate teletraffic models for mobile networks, such as those that support cellular or Personal Communications Services (PCS). The movement of the user implies that existing teletraffic models for the Public Switched Telephone Network (PSTN) cannot be used for mobile networks. Most mobile communications researchers simplify analysis by assuming a temporal and spatial uniform distribution of mobile calls. This could lead to wrong conclusions about real mobile networks (Lam *et al.*, 1997). From the foregoing, it is clear that previous studies on the subject matter were either carried out at the radio interface (cell level) on a mobile network or at the A or E-interface on a fixed network. However, with pulse code modulation congestion (PCMC) on a BS - MSC link or MSC - MSC link, the call would be blocked. This is true even if the radio interface is congestion-free. For a zero blocking probability, all the links between the MS and the MSC should be devoid of congestion. This present study seeks to address this missing link. Our attempt is to utilize the training capability of LMA to develop a model to predict, and thus control, telephone traffic congestion, more particularly PCMC, in a mobile network using artificial neural networks.

Artificial Neural Networks : Artificial neural networks (ANN) are mathematical tools originally inspired by the way human brain processes information (Hippert *et al.*, 2001). They are intelligent systems that are related in some way to a simplified biological model of the human brain. ANN are essentially function approximators that transform inputs into outputs to the best of their ability. They are composed of many simple elements, called neurons, operating in parallel and connected to each other in the forward path by some multipliers called the connection weights.The Levenberg-Marquardt Algorithm (LMA) is one of the variants of the basic backpropagation algorithm. This algorithm has been shown to be the fastest method for training moderate-sized feed-forward neural networks (up to several hundred weights). It also has an efficient implementation in MATLAB® software, since the solution of the matrix equation is a built-in function, so its attributes become even more pronounced in a MATLAB® environment (Demuth and Beale, 2008). The LMA uses this approximation to the Hessian matrix (**H**) in the following Newton-like update:

$$\mathbf{X}_{k+1} = \mathbf{X}_k - [\mathbf{J}^T\mathbf{J} + \mu\mathbf{I}]^{-1}\, \mathbf{J}^T\mathbf{e} \qquad \ldots (1)$$

$\mathbf{X}_k$ is a vector of current weights and biases, **J** is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, **e** is a vector of network errors, $\boldsymbol{\mu}$ is a scalar, **I** is the identity matrix and $\mathbf{J}^T$ is the transpose of **J**. The expressions for **J** and **H** are given as:

$$\mathbf{H}(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \qquad \ldots (2)$$

$$\mathbf{J} \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \qquad \ldots (3)$$

II. MATERIAL AND METHOD

The experimental procedure involves data collection and pre-processing, network modeling and training and simulation.

Data Collection and Pre-processing : Twelve month hourly traffic data retrieved from the automated traffic recordings of a MSC located in Bauchi metropolis was used. The various counters of the traffic data were careful selected by intuition so as to take those that actually affect congestion. Some of the variables selected include: Time of the day, Total call duration (in seconds), Peer office congestion, Seizure traffic (in Erlangs), Seizure attempts, Answer times, No answer after ringing times and Called subscriber busy times. Twenty four observations were made every day. Over the twelve month period, a total of 8,760 observations were made and extracted as data samples.The data were then preprocessed each time before being introduced to the Neural Network for training. Pre-processing involves scaling all the data in the range (-1, 1). This improves the accuracy of the neural network training.

Proposed Network Model and Training :A Multi-Layer Perceptron Feed-forward neural network with Levenberg-Marquardt Algorithm was used to train the network model. Training involves varying the network weights and biases and allowing the network to learn the different knowledge about the inputs and outputs. The number of neurons in the hidden layer (n) was varied to achieve optimum performance of the network. After a few experimental run, the number of neurons in the hidden layer was set at fifteen. The input parameters into the network include: Time of the day, Seizure attempts, Answer times, No answer after ringing times, Called party busy times, Seizure traffic (in Erlangs) and Total call duration (in seconds) as shown in figure 2. The output parameter is the peer office congestion times which measure the level of congestion of the trunk group. Equations (4) and (5) represent the input and output functions with their variables respectively.

$$I = f(T, S_a, A, N, C_b, S_t, T_c) \qquad \ldots (4)$$

$$Y = f(C) \qquad \ldots (5)$$

I = Input; Y = Output; T = Time of the day; S_a = Seizure attempts; A = Answer times
N = No answer after ringing times; C_b = Called subscriber busy times; S_t = Seizure traffic
T_c = Total call duration

Once the network weights and biases have been initialized, the network is ready for training. It is trained for function approximation and pattern association. A Mean Squared Error (MSE) of 10^{-4} was set as the error goal and the initial value of µ was set at 10^{-3}.

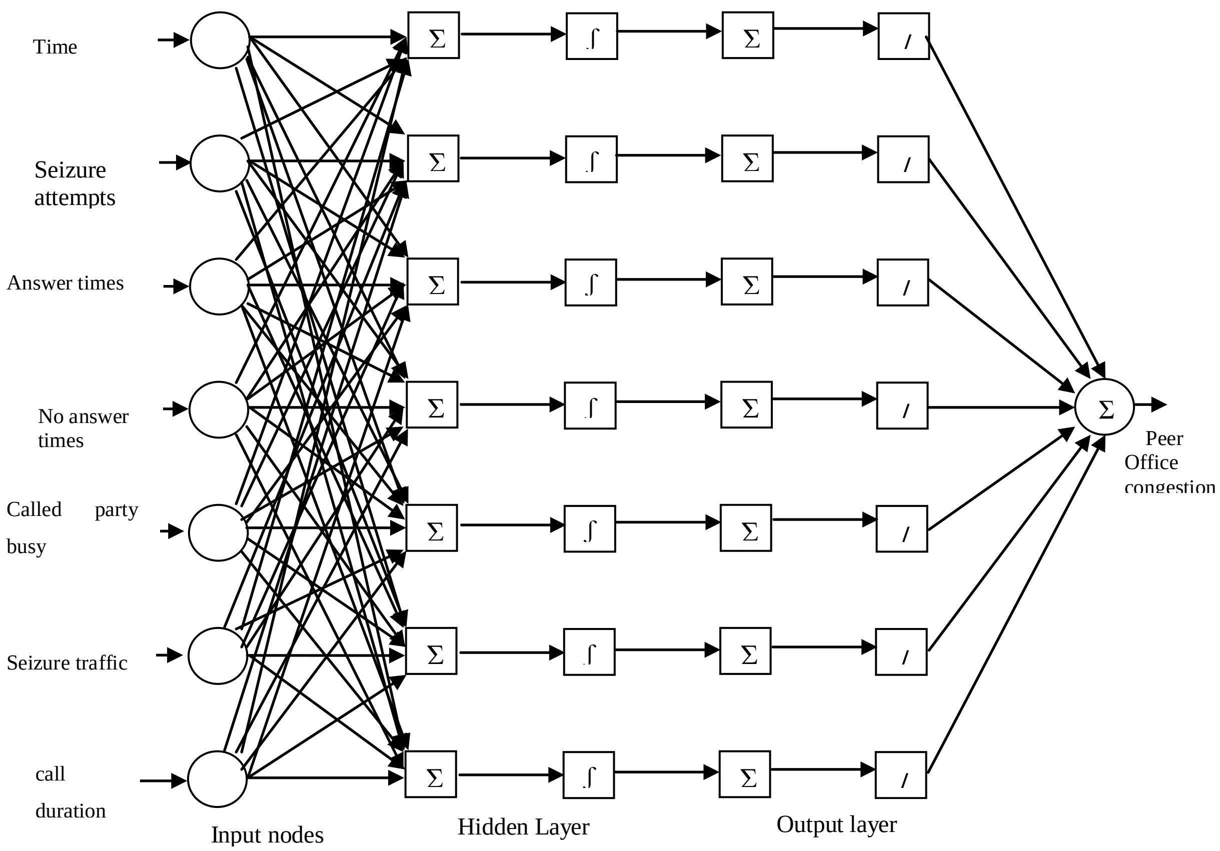

Figure 2: The two-layer feed-forward neural network

Simulation : After proper training of the network, simulation follows. Simulation involves presenting new set of inputs to the network and allowing the network to produce an output. After simulation, regression analysis was performed to determine the correlation between simulated outputs and targets.

III. RESULTS

A typical training session is shown in figure 3. It can be seen that this training instance took 0.04 seconds to complete 18 iterations. The error was minimized at 4.56 x 10^{-3}, very close to the set goal. The best validation performance of 1.4345 x 10^{-3} was achieved after twelve iterations.Figure 4 is the performance training window, which shows a plot of the training errors, validation errors, and test errors. It gives small mean-square error, no significant over-fitting and similar validation set errors and test set errors characteristics.

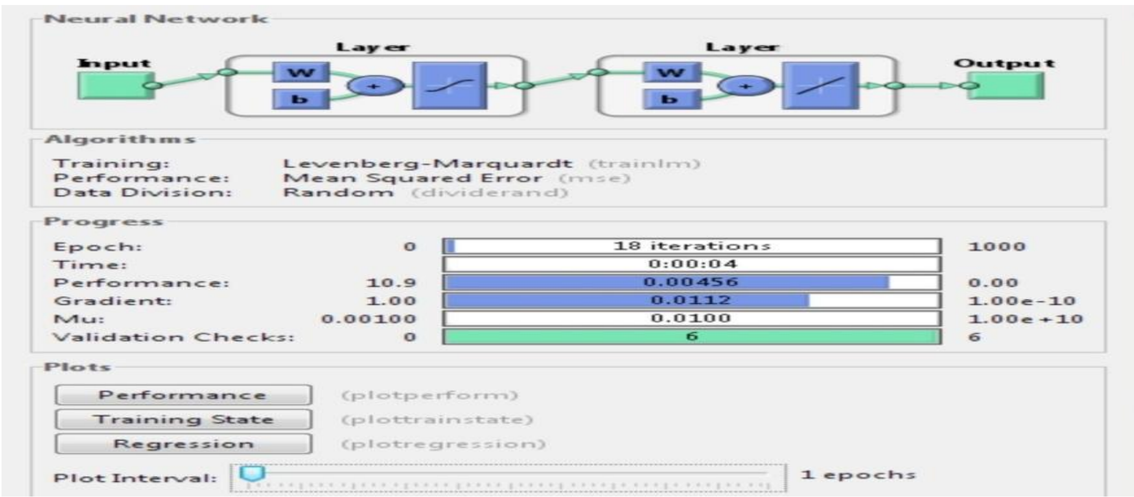

Figure 3. A typical training session using neural network tool

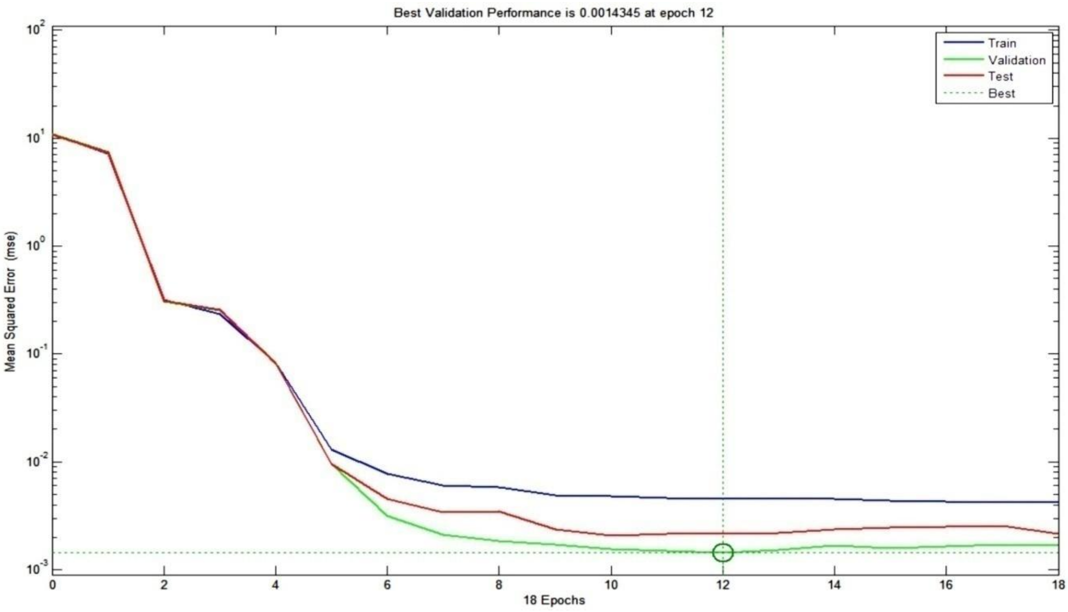

Figure 4. A plot of training, validation and test errors

Shown in figures 5(a)-(c) are prediction comparisons for three different days. It is clear that the simulated congestion volumes tracts the actual congestion volumes well.

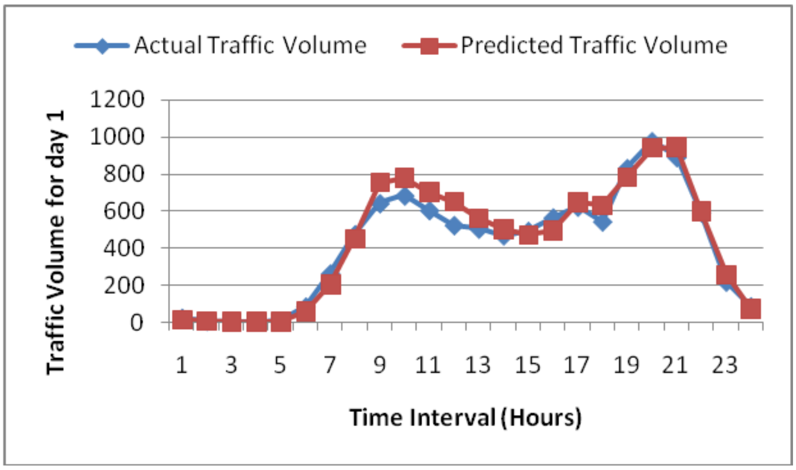

Figure 5(a). Actual congestion vs. predicted congestion (day 1)

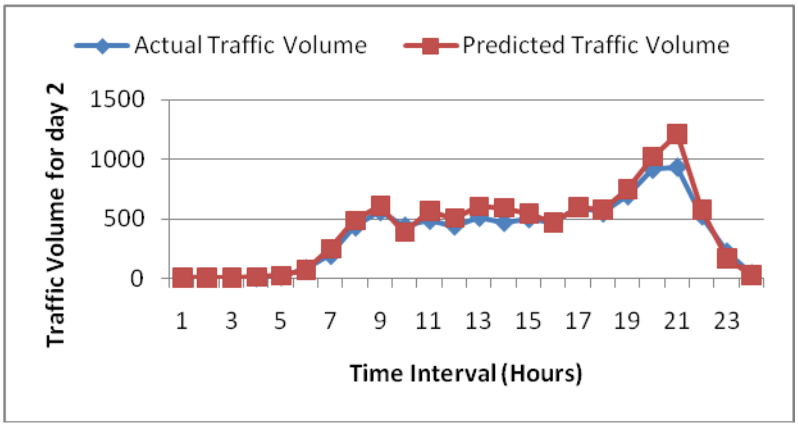

Figure 5(b). Actual congestion vs. predicted congestion (day 2)

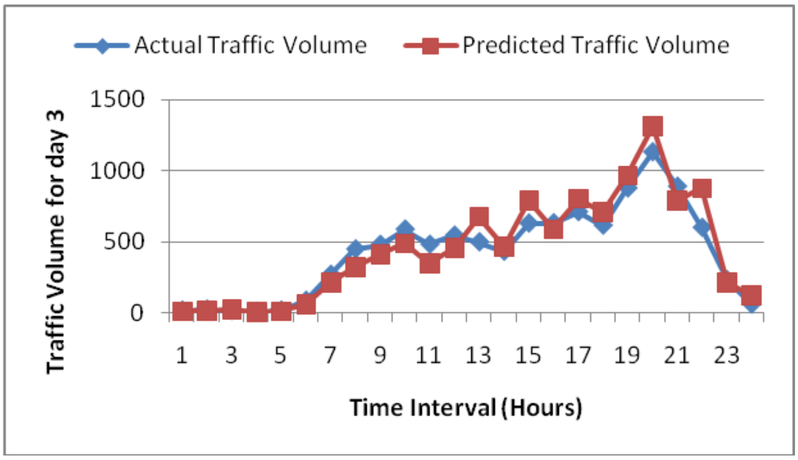

Figure 5(c). Actual congestion vs. predicted congestion (day 3)

Regression Analysis : To further evaluate the performance of the developed neural network model, a linear regression between the network outputs and the corresponding targets was carried out. Three parameters were returned. The first two, m and b, correspond to the slope and y-intercept of the best linear regression relating targets to network outputs respectively. A perfect fit gives a slope of 1 and zero intercept on the y-axis. The third variable returned (R-value) is the correlation coefficient between the outputs and targets. An R-value of 1 indicates perfect correlation between the targets and outputs. Figure 6 shows the regression plot carried out on the first day prediction. The network outputs were plotted against the targets as open circles. The best linear fit is indicated by a solid line while the perfect fit is indicated by the dash line. A slope of 0.93, an intercept of 10 and an R-value of 0.986 were obtained. This indicates a close relationship and high correlation between the actual outputs and the predicted outputs.

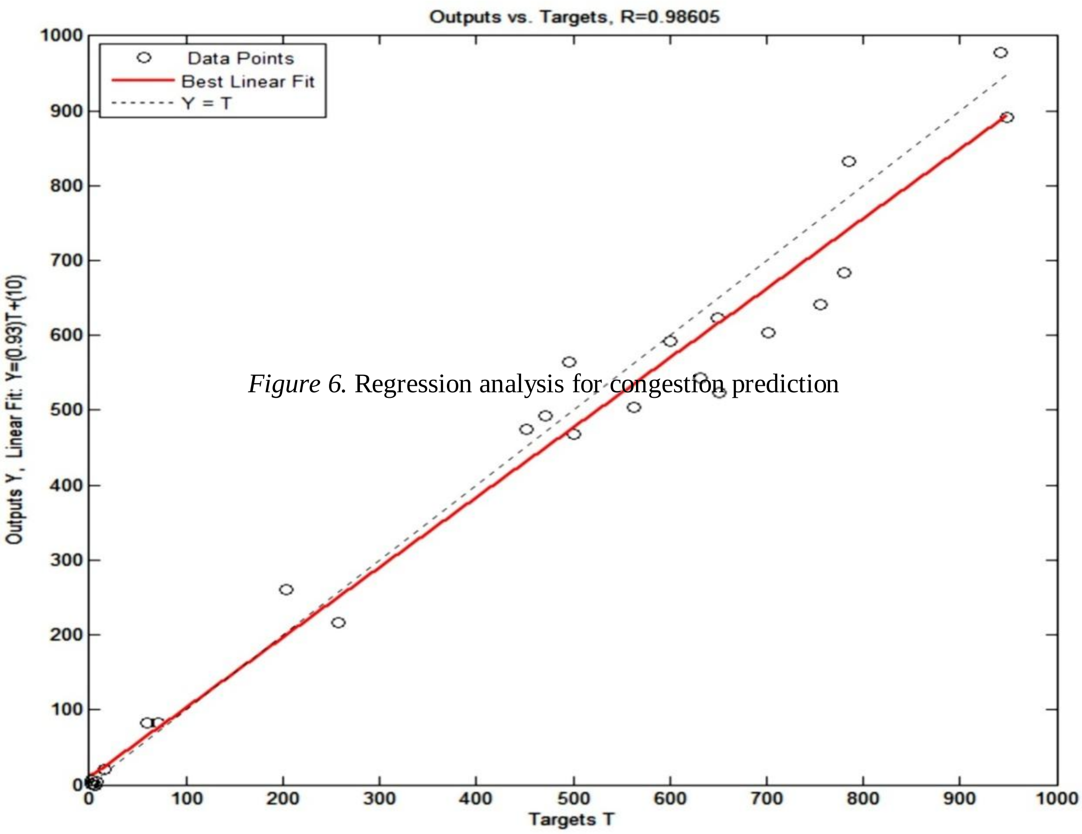

Figure 6. Regression analysis for congestion prediction

Figure 6. Regression analysis for congestion prediction

IV. CONCLUSION

Telephone traffic congestion pattern of a mobile network was studied. The multilayer perceptron feed-forward neural network was used to model the network. The network was trained using actual traffic data. The trained network was simulated with new input data and our results show that the behavior of the developed model is very close to a real network situation. With availability of relevant historical traffic data, artificial neural networks can model the behavior of mobile network to predict the occurrence of network irregularities.

REFERENCES

[1] Alarape, M.A.; Akinwale, A.T. and Folorunso, O. (2011). A Combined Scheme for Controlling GSM Network Calls Congestion. *International Journal of Computer Application*, **14**(3), *pp*: 47-53.

[2] Andreas, P. and Ahmet, S. (1999). *Congestion Control: Book CI Congestion-Control-Final.* (Tech. Doc.), University of Cyrus, Nicosia Cyrus.

[3] Demuth, H. and Beale, M. (2008). **Matlab Neural Network Toolbox Help.** Version R2008a, The MathWorks, Inc.

[4] Harte, L.; Levine, R. and Livingston, G. (1999). **GSM Superphones.** McGraw Hill **71**: 45-47.

[5] Hippert, H.; Pedreira, C. and Souza, R. (2001). Neural Networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Neural Networks,* **16***(1)*: 44–55.

[6] ITU-T Recommendation 1.371, (1993). **Traffic Control and Congestion Control in B-ISDN.**

[7] Kuboye, B.M. (2010). Optimization models for minimizing congestion in Global System for Mobile Communications (GSM) in Nigeria. *Journal of Media and Communication Studies* **2**(5), *pp*: 122-126.

[8] Kuboye, B.M. (2006). *Development of a framework for managing of congestion in GSM in Nigeria*. Unpublished master's thesis, department of electrical and electronics engineering, University of Portharcourt, Nigeria.

[9] Lam, D.; Cox, D.C. and Widom, J. (1997). Teletraffic Modeling for Personal Communications Services. *IEEE Communications Magazine,* February 1997, Pp. 79-87.

[10] Markus, E.D.; Okereke, O.U. and Agee, J.T. (2011). Predicting Telephone Traffic Congestion using Multilayer Feed-forward Neural Networks. *Journal of Advanced Materials Research,* **367**:191-198.

[11] Mehrotra, A. (1997). **GSM System Engineering.** Artech house, Inc., Pp: 70-73

[12] Ojesanmi, O.A.; Oyebisi, T.O.; Oyebode, E.O. and Makinde, O.E., (2011). Performance Analysis of Congestion Control Scheme for Mobile Communication Network, *International Journal of Computer Science and Telecommunications,* **2**(8).