International Journal of Engineering Research and General Science Volume 3, Issue 2, March-April, 2015
ISSN 2091-2730

Time Series Forecasting of Packet Loss Rate Using Artificial Neural Network Based on Particle Swarm Optimization

[1]Manish P Ganvir, [2]Dr. S.S.Salankar

[1]M.Tech Student, Department of Electronics & Telecommunication

G.H. Raisoni College of Engineering, Nagpur-16

E-mail- manishganvir9833@gmail.com

Contact No - +91-9028897951

Abstract— Packet loss severely degrades the quality of service of multimedia communication in a Wi-Fi network. In this paper, a time-series prediction model for the packet loss rate (PLR) is developed. The reason for prediction of PLR is very much useful in congestion control mechanisms. An accurate prediction method would therefore be very helpful. An artificial neural network is used as a prediction model and it is trained with Particle swarm Optimization (PSO) as a training algorithm in order to get accurate prediction of packet loss rate.

Keywords— Artificial Neural Network, packet loss rate, Particle Swarm Optimization, Time series forecasting.

INTRODUCTION

The quality of service is severely degraded by the packet loss multimedia applications. Packet loss generally occurs during the time of heavy congestion as nodes gets overburdened with packets and hence some packets have to be discarded. For that reason, any system that will help in reducing the Packet Loss will be very much useful. There are several reasons which leads to predicting the packet loss rate. As the TCP protocol is based on a complex retransmission algorithm which is not appropriate for the real time applications, hence UDP-based transmission are used for such applications. The main difference between UDP based traffic and TCP based Traffic is that if congestion occurs then TCP will reduce the traffic rate but UDP don't have such provision. That will not only worsen the congestion but also violates the fairness issue. Hence UDP traffic should also use the traffic rate control mechanisms as it is used by the TCP traffic. The two entities that define the rate adjustment mechanism is Packet Loss rate (PLR) and round-trip time (RTT). RTT can be defined as the length of time the signal takes to reach to the receiver and acknowledged back by the receiver. So a new approach is to predict PLR instead of using the earlier measured values. Rather than using the reactive approach a predictive approach will be faster to trace congestion problems. The other reason for predicting packet loss rate is UDP based real time multimedia traffic in which extra packets are added using Forward error correction mechanism to recover lost packets. This leads to excessive bandwidth usage. Hence it is imperative to send only the required amount of packets. An accurate prediction of packets will be handful in such conditions.

Artificial Neural networks are a type of nonlinear systems competent of learning and performing tasks achieved by other systems. Neural network systems are robust as even the small occurrence of errors does not interfere with the proper operation of the system having neural network building blocks. This feature of the neural networks makes them rather appropriate for the prediction task. Artificial neural network is a kind of machine learning approach which models human brain and consists of a number of artificial neurons that are linked together according to a specific network architecture. The main function of the neural network is to transform the inputs into meaningful outputs.These systems have the potential to capture highly non-linear mappings between input and output.

TIME SERIES FORECASTING

A time series is defined as a collection of data recorded over a period of time—weekly, monthly, quarterly, or yearly.A time series is an order which is measured at successive points in time spaced at a uniform time interval. Time series analysis consists methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. It is a stochastic process which has sequence of observations of a random variable. For example the monthly demand for a product, the annual sale of a product and the daily volume of the flow of the river. The foundation for decision models is provided by the forecasting of time series data in operations research. Time series analysis gives tools for choosing a model that can be useful in forecasting future events. Modelling of a time series is a type of statistical problem. The parameters of a model are estimated with the help of forecast in computational procedures. These models work on the assumption that probability distribution varies with the observations within an underlying function of time. There exist many models used for time series, however, there are three very broad classes that are used most often. These are the autoregressive (AR) models, the integrated (I) models, and the moving average (MA) models. These models are often combined to form a new model. For example, the autoregressive moving average model (ARMA) combines the (AR) model and the (MA) model. Another example of this is the autoregressive integrated moving average (ARIMA) model, which combine all three of the models previously mentioned. Autoregressive (AR) models are the mostly used model for time series data. The autoregressive process is basically a difference equation determined by random variables. The key component in modelling time series is the distribution of such random variables.

FORECASTING MODELS

1. Autoregressive (AR) Model

A model which depends only on the previous outputs of the system is called an autoregressive model (AR).One of the most important consideration is the choice of the number of terms in the AR model, this is known as its order p. The AR-model of a random process in discrete time is defined by the following expression:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots \alpha_p X_{t-p} + Z_t$$

The model parameters are found by solving a set of linear equation obtained by minimizing the mean squared error. The characteristic of this error is that it decreases as the order of the AR model is increased.

2. ARMA Model

ARMA models combine auto regressive (AR) and moving *average* (MA) models. AR models are a pole-only model:

$$H(z) = \frac{1}{1+a_1z^{-1}+a_2z^{-2}+\cdot\ \cdot\ \cdot+a_pz^{-p}}$$

where p is the model order and $a_1,a_2,\ldots,a_p$ are the model coefficients. MA models are zero-only models:

$$H(z) = \frac{b_0+b_1z^{-1}+b_2z^{-2}+\cdot\ \cdot\ \cdot+b_qz^{-q}}{1}$$

where q is the model order and $b_0,b_1,\ldots,b_q$ are the model coefficients. Although both models are based on time domain samples and can be expressed as sample-by-sample algorithms in the time domain, they possess very important spectral properties. AR models have the ability to estimate the *power spectral density* (PSD) of processes whose spectra contain sharp peaks and broad valleys, while MA models can estimate the PSD of processes whose spectra contains sharp valleys and broad peaks.

An ARMA model has the form:

$$H(z) = b_0+b_1z^{-1}+b_2z^{-2}+\cdot\ \cdot\ \cdot+bqz-q$$

$$1+a_1z^{-1}+a_2z^{-2}+\cdots+a_pz^{-p}$$

where the model denominator has order p and coefficients $a_1, a_2, \ldots, a_p$ and the numerator has order q and coefficients

$b_0, b_1, \ldots, b_q$.

3. ARIMA Model

Auto Regressive Integrated Moving-Average (ARIMA) is another important forecasting approach, going over model validation, parameter estimation, and model identification. The main advantage of this method is that it relies on the accuracy over a wider domain of series, despite being more complex, in terms of usability and computational effort, than Holt-Winters. The global model is based on a linear combination of past values (AR components) and errors (MA components), being named Auto Regressive Integrated Moving-Average (ARIMA).

The non seasonal model is denoted by the form ARIMA(p; d; q) and is defined by the equation:

$$\Phi_p(L)(1 - L)^d y_t = \theta_q(L)e_t$$

where y_t is the series; e_t is the error; L is the lag or backshift operator ;

$\Phi_p = 1 - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p Lp$ is the AR polynomial of order p; d is the differencing order; and

$\theta_q = 1 - \theta_1 L - \theta_2 L^2 - \ldots \theta_q L_q$ is the MA polynomial of order q.

When the series has a non zero average through time, the model may also contemplate a constant term in the right side of the equation. To create multi-step predictions, the one step-ahead forecasts are used iteratively as inputs.

ARTIFICIAL NEURAL NETWORKS

An Artificial neural network can be defined as a processing device, or an algorithm. ANNs have three great advantages over traditional methods. At first, they have universal approximation capabilities second, they can recognize ‘‘on their own’’implicit dependencies and relationships in data third, they can ‘‘learn’’ to adapt their behavior viz., prediction, to changed conditions quickly and without complication. Neural models are innate candidates for forecasting due to their nonlinear and noise tolerance capabilities. The basic idea is to train a NN with past data and then use this network to predict future values. The use of NNs for TSF began in the late eighties with encouraging results and the field has been consistently growing since. Although different types of NNs have been applied in the forecasting literature (e.g. Recurrent Networks , the majority of the studies uses the Multilayer Perceptron network).

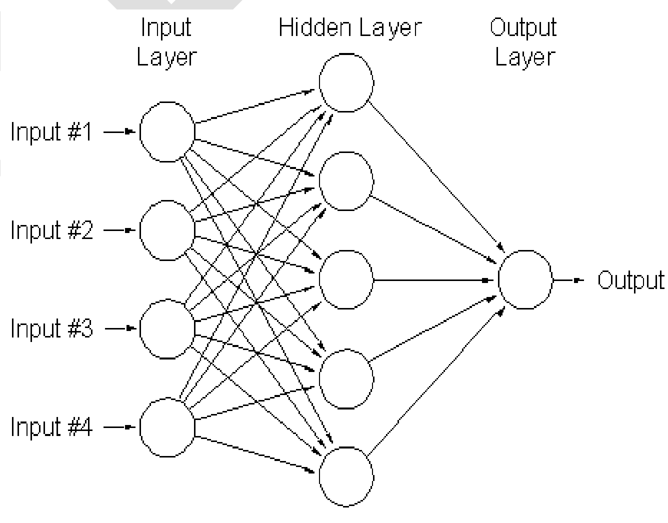


Fig. 1. The Multilayer Perceptron architecture

These architectures have a group of neurons -the input layer-, which are fed by external stimuli. The input units send these stimuli to hidden neurons, which are combined in one or more internal layers. These hidden units process the information they receive, and forwrd their results to the last layer of neurons i.e the output layer. The neurons in the output layer gives the final ouput. The layers are connected through information links, whose weights have to be determined in order to relate desired outputs to inputs. The computations carried out inside each neuron refers to: i) performing the weighted average of its impinging inputs, ii) sending this average through a activation function, and iii) forwarding the activation function output to the next layer of neurons. These calculations are generally performed simultaneously by all the neurons in a given layer, so that the response delay depends on the number of layers. The process of adjusting weights and biases is known as network training, and the algorithm which performs this task is the so called backpropagation algorithm. It essentially consists of a gradient-descent algorithm to reduce the error between actual and desired network outputs by modifiying weights and biases going backward from the output-layer to the input-layer connections. The other type of learning algorithms that can be used are Genetic algorithm and Particle Swarm Optimization (PSO) algorithm.

PARTICLE SWARM OPTIMIZATION (PSO)

The PSO is based on the behavior of colony of living things. PSO is based on the behavior of colony or a swarm of insects such as ants, bees, termites, wasps, and a flock of birds or school of fish.It mimics the behavior of social organism. It is a population based algorithm. The word “Particle” denotes a bird in a flock or bee in a colony.” Swarm” means moving particles which have certain velocity.”Optimization” means obtaining best results from given circumstances. The PSO algorithm was originally proposed by Kennedy and Eberhart in 1995.They proposed an algorithm where each particle is located randomly in space. Particle is assumed to have two characteristics: a) Position b) Velocity. Each particle wonders around in the space and remembers its best position .This individual best position (obtained by using its own knowledge) is called “Pbest”. Particle achieve best position in a group (obtained by sharing knowledge among a group) is called “Gbest”. The formulae used to find modified position and velocity are shown in equation (1) and (2)

$$Xi(t) = Xi(t-1) + Vi(t) \tag{1}$$

$$\begin{aligned} Vi(t) &= w * Vi(t-1) + \Phi 1 * rnd1 * (Pi - Xi(t-1)) \\ &+ \Phi 2 * rnd2 * (Pg - Xi(t-1)) \end{aligned} \tag{2}$$

$$Vi(t) = Inertia + Cognitive + Social. \tag{3}$$

Where, Xi(t) = New particle position

Xi(t-1) = Previous position

Vi(t) = New particle Velocity

Vi(t-1) = Previous Velocity

W = Inertia Weight

Φ1 & Φ2 = Two positive numbers

rnd1&rnd2 = Two random numbers with uniform distribution in the range of (0,1)

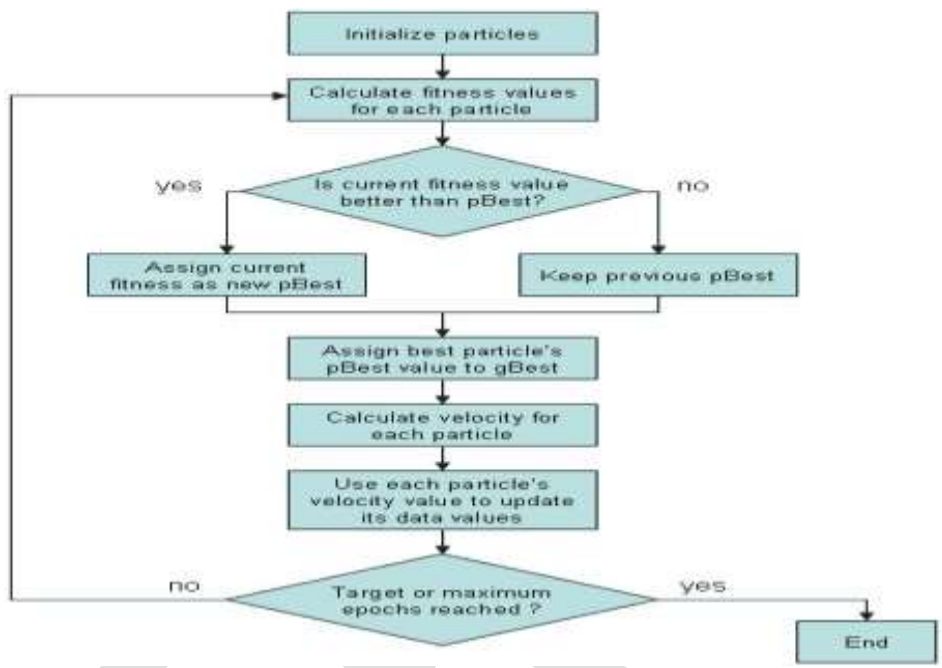
Pi = Individual best position (Pbest)

Pg = Global best position (Gbest)

Equation (3) shows three components.

First component shows the term inertia which develop the tendency of the particle to continue in the same direction in which it was travelling.Second component shows the linear attraction towards the best position found by the given particle. This component is referred to “self knowledge”.Third component shows linear attraction towards the position found by any particle .This component is referred to “group knowledge.

PSO FLOWCHART



PSO ALGORITHM

Consider a objective function which has to maximize or minimize. Suppose Maximize, Take maximizing function to be f(x).

With $X^l \leq X \leq X^u$

Where $X^l \longrightarrow$ Lower bounds of X

$X^u \longrightarrow$ Upper bounds of X

The PSO can be applied through the following steps:

1. Assume Size of the swarm (number of particles) is N.

2. Generate the initial position of X in the range X^l and X^u randomly as $X_1,X_2,- - - - - - X_N$.

Particle position ‘j’ in iteration ‘i’is given by X^i_j .Initially particles are having values $X_1(0),X_2(0),- - - - - - X_N(0)$.

And the objective function is given by $f_1(0),f_2(0)- - - f_N(0)$.

3. Set iteration number as i =1.

4. a) i)Find Pbest with highest value of objective

Finction for j^{th} particle.

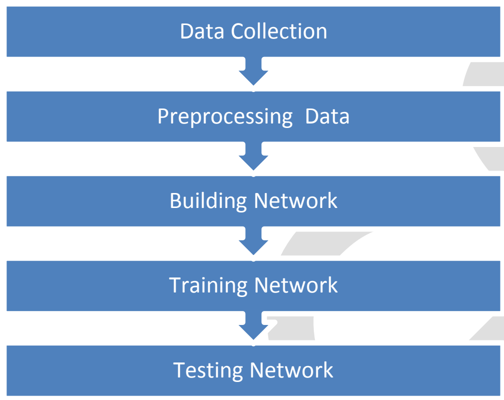
ii)Find Gbest with highest value of objective function for any particle in N number of particles.

b)Find Velocity of particle j in i_{th} iteration using equation(2)

c)Find position of particle j in i_{th} iteration using equation(1)

5. Check the convergence of current solution. If position of all particle converges to same set of values stop iteration. Unless repeat step 4 by updating equation number as i = i+1.And computing new values of Pbest and Gbest. The process is continued until all particles converge to same optimum solution.

DESIGNING ANN MODELS



Designing ANN models follows a systematic procedure which includes collection of data from a database then preprocessing the data, building the network and training the network with a training algorithm and finally testing the network for accurate predictions.

DATABASE COLLECTION

The database is collected from the open source site http://crawdad.org where the experiment performed by researchers at outdoor rural measurement campaign using two IBM Thinkpad R40e laptops (Celeron 2 GHz with 256 MB ram running Debian Linux with a 2.6.8 kernel), equipped with CNet CNWLC-811 IEEE 802.11b PCMCIA wireless cards and standard drivers. The rural environment was a wide uncultivated field with an unobstructed line of sight, far from buildings, cell phone antennas and power lines.The database contains 1000 samples of packet loss rate.

PREDICTION ANALYSIS AND RESULT COMPARISON

A time series of 1000 samples is used as a database to train the MLP neural network using different training algorithms. The 700 samples are used for training the network.150 samples are used for testing the network and the remaining 150 samples used to validate the neural network model The training process stops when the error between the predicted value and actual value is minimized. The MLP neural network used contains one input layer, one hidden layer and one output layer with sigmoidal activation function in the hidden layer and linear activation function in the output layer. The MLP network is trained with different training algorithms and compared with respect to there mean square error values. Mean square error is defined as the difference between the predicted value and actual value.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2.$$

Type of Algorithm	*MSE*
PSO	*0.04*
LM	*0.15*
SCG	*0.28*

Table 1. Comparison of different training algorithms

Table I summarizes the prediction results on input samples for the proposed methods as well as the competing methods. The notation PSO means Particle swarm Optimization. LM means Lavernberg-Marquardt Algorithm. SCG means Scaled Conjugate Gradient Algorithm. MLP trained with PSO gives least mean square error as compared to other algorithms.

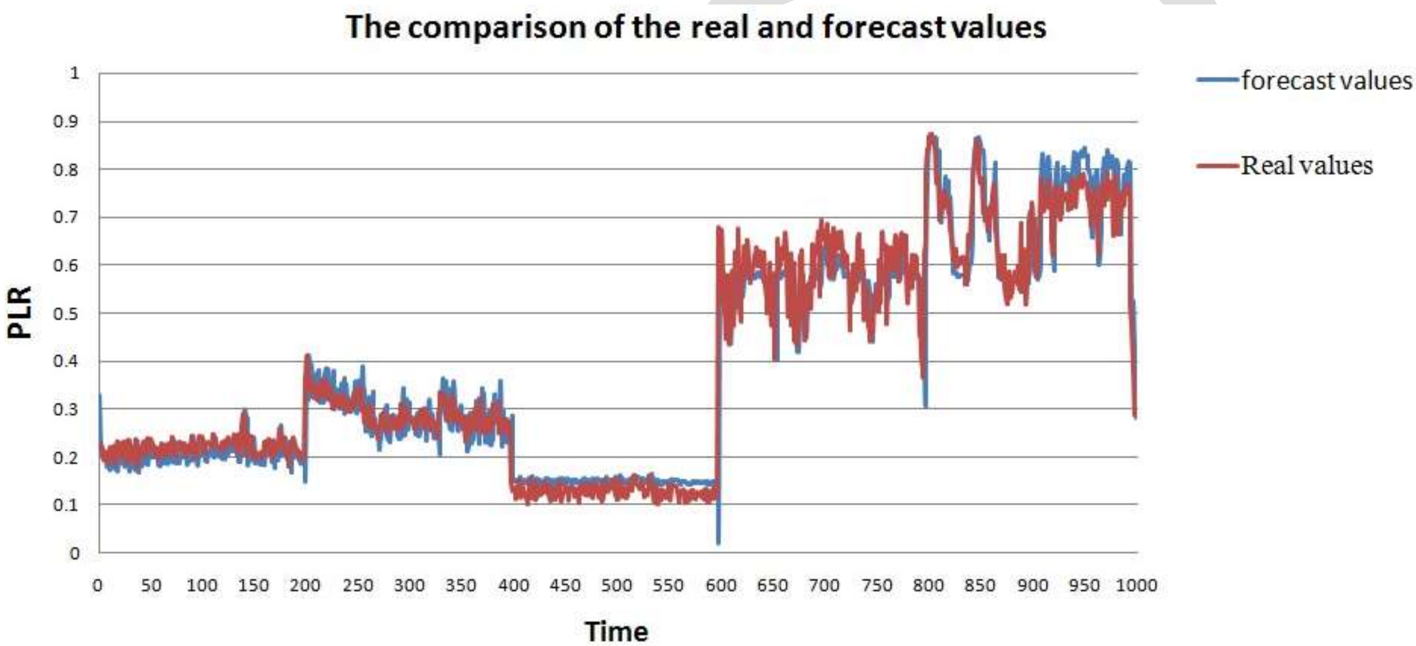


Fig. 3. Prediction and actual time series for the input samples for the PSO-ANN model. The time axis unit is "sample number," while the PLR axis unit is lost packets.

CONCLUSION

In this paper, the problem of PLR prediction is considered. The quality of real-time multimedia traffic can be improved by accurate prediction thereby reducing the congestion. Several neural network models for Packet Loss Rate forecasting are studied in this work. The neural network model when trained with Particle Swarm Optimization gives good prediction accuracy as compare to other algorithms. In addition, it is faster than the other methods. According to the discussion and the comparison of model forecast accuracy shows that Particle Swarm Optimized Artificial Neural Network is the best model for PLR forecasting. This type of network can be very efficient in terms of predicting future values.

REFERENCES:

[1] S. B. Moon, "Measurement and analysis of end-to-end delay and loss in the Internet," Ph.D. dissertation, Dept. Comput. Sci., Univ. Massachussets, Amherst, MA, 2000.

[2] Amir F. Atiya, Sung Goo Yoo, Kil To Chong, and Hyongsuk Kim, "Packet Loss Rate Prediction Using the Sparse Basis Prediction Model " IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 18, NO. 3 ,2007

[3] Y.-C. Su, C.-S. Yang, and C.-W. Lee, "The analysis of packet loss prediction for Gilbert-model with loss rate uplink," in *Inf. Process. Lett.*, 2004, vol. 90, pp. 155–159.

[4] J.Widmer, R. Denda, and M. Mauve, "A survey on TCP-friendly congestion control," *IEEE Network*, vol. 3, pp. 28–37, 2001.

[5] P. S. Rossi, G. Romano, F. Palmieri, and G. Ianello, "Joint end-to-end loss-delay hidden Markov model for periodic UDP traffic over the Internet," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 530–541, Feb. 2006.

[6] L. Roychoudhuri and E. Al-Shaer, "Real-time packet loss prediction based on end-to-end delay variation," *IEEE Trans. Network Service Manager.* , vol. 2, no. 1, Nov. 2005.

[7] S. G. Yoo, K. T. Chong, and S. Y. Yi, "Neural network modeling of transmission rate control factor for multimedia transmission using the Internet," in *Lecture Notes in Computer Science*, ser. 3399. Berlin, Germany: Springer-Verlag, 2005, pp. 851–862.

[8] H. Mehrvar and M. R. Soleymani, "Packet loss prediction using a universal indicator of traffic," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Helsinki, Finland, Jun. 11–14, 2001, vol. 3, pp. 647–653.

[9] D. Ashok kumar, S. Murugan "Performance Analysis of Indian Stock Market Index using Neural Network Time Series Model" Pattern Recognition, Informatics and Mobile Engineering (PRIME) February 2013.

[10] Ignacio Fernandez Anitzine, Juan Antonio Romo Argota,Fernado Perez Fontan, "Influence of Training set selection in Artificial neural Network based Propagation Path Loss Predictions" International Journal of Antennas and Propagation,vol-12,Article ID 351487,2012

[11] Yan Zhu, Guanghua Zhang, and Jing Qiu "Network Traffic Prediction based on Particle Swarm BP Neural Network" Journal Of Networks, Vol. 8, No. 11, November 2013.

[12] Yogesh D. Barve "Neural Network Approach to the Prediction of Percentage Data Packet Loss for Wireless Sensor Networks" 41st Southeastern Symposium on System Theory. 2009.

[13] Ratnadip Adhikari, R. K. Agrawal, Laxmi Kant, "PSO based Neural Networks vs. Traditional Statistical Models for Seasonal Time Series Forecasting" IEEE International Advanced Computing Conference (IACC), 2013

[14] Sapna Katiyar "A Comparative Study of Genetic Algorithm and the Particle Swarm Optimization" *AKGEC International Journal of Technology*, vol. 2.