International Journal of Computer Science and Applications

Vol. 14 No. 2, pp. 130 - 141, 2017

Intelligent Time Series Model to Predict Bandwidth Utilization

Theyazn H. H. Aldhyani
School of Computer Sciences
North Maharashtra University
Jalgaon (M.S), India.
Email:th0ha0@yahoo.com

Manish R. Joshi
School of Computer Sciences
North Maharashtra University
Jalgaon (M.S), India
joshmanish@gmail.com

What amount of data shall be transmitted in future on particular network? This is the issue in the brain of a network administrator and network engineers while designing the network and for reserving the required bandwidth. In order to enhance the efficiency of the resource utilization and to maintain the bandwidth is very challenging issue in network traffic.
In this paper, an intelligent model is proposed that combines the clustering with Holt-Trend Exponential Smoothing (HTES) time series prediction model to predict expected bandwidth utilization on wide area network. The novelty of the proposed model is integrated results that obtained from Holt-Trend Exponential Smoothing model time series prediction model with centriods that are obtained from clustering approach.
The model is evaluated using real network that is collected from Lawrence Berkeley Laboratory. Standard metrics like MSE, NMSE, RMSE, MAE and MAPE are used to evaluate the results of proposed model.
Evaluation and comparison of the proposed model with existing models is presented. Our model outperforms of all the existing models.

Keywords: Network traffic prediction, time series model, k-means.

1. Introduction

With the increasing volume of big data these applications have imposed significant challenges on users to achieve the finest possible network performance on Internet. Currently, Peer-to-Peer (P2P) network and applications capture a large share of the total bandwidth on the Internet. Due to the increased the volume of traffic which are transmitted on geographically distributed, the data volume is continuously growing of it it's network speed and bandwidth. Therefore, the network performance is influenced by the increase in the bandwidth consumption [J. Park *et al.* (2015)]. If the traffic growth goes beyond the certain limit of its total capacity, the network will become quite unstable. The prediction of bandwidth utilization is playing very important role, it is valuable for E-commerce as well as it is a prerequisite for all users who are connected with each other and who need the best performance on speed of Internet for data exchange. In 1992 the total Internet traffic per day was 100 GB, rapidly after one decade later in 2002 it has skipped upto 100 GB per second, even as in 2014 it has gone up to 16,144 GB per second. The predication for 2019 is expected to be more than 51,974 GB per second [cisco (2016)], [Man-Soo Han *et al.* (2014)]. Therefore, numbers of researchers have proposed various network traffic prediction models and traffic engineering mechanisms [Bin Yang *et al.* (2016)], [Samar Raza *et al.* (2016)], [Tao Peng *et al.* (2015)], [Lun Zhang *et al.* (2013)], [Poo KuanHoong *et al.* (2012)] to predict the growth of bandwidth requirement in networks.

In this paper, we proposed intelligent model to enhance network traffic accuracy for bandwidth utilization prediction. The novelty of the proposed model is used of clustering technique to enhance conventional HTES model for short-term network traffic flow prediction. Similarity of network traffic data series is aggregated with time intervals 0.1s for reducing the length of actual traffic data set. Finlay, it is observed that the proposed model has optimized the prediction ratio.

The rest of paper is organized as follows. Section 2 discusses related work. The description of methodology of research work is discussed in section 3. Section 4 presents the experimental analysis followed by results analysis. Finally, section 5 concludes the paper.

2. Related Work

Several report and research outcomes of network traffic prediction published in literature review are interested in solving the problem for improving the efficiency and effectiveness of network traffic monitoring by predictions data packet flow in advance. Therefore, an accurate traffic prediction model should have the ability to capture the prominent traffic characteristics, e.g. short-range and long-range dependences, self-similarity. In the following sections we will discuss some of the works done so far, which has motivated this work.

[Yanhua *et al.* (2013)] proposed AutoRegressive Integrated Moving Average (ARIMA) linear time series model for prediction of network traffic. Authors experimented with mobile of network from Heilongjiang chain at different time interval. They used Mean Absolute Percentage Error (MAPE) metric to measure and evaluate prediction results of ARIMA approach. They applied correlation coefficient function to represent time series. According to authors, the ARIMA approach achieved high prediction. [**?**] described linear ARIMA model for modeling and predicting wireless network traffic. The authors used GSM wireless mobile real data from Tianjin at china mobile of different time scales. They applied Minimum Mean Square Error (MMSE) function to measure and determine prediction performance of ARIMA model. A comparative prediction between actual data and prediction data is presented. They concluded that the ARIMA technique obtained better prediction. They observed relative error between prediction values and original values are less than 0.02. [Jian Kuan *et al.* (2013)] proposed hybrid model two dimensional correction and Single Exponential Smoothing (SES) for prediction the mobile network. They collected data from Xinjiang mobile company on different time interval days and hours. Their experiments shows that their model has performed well, a comparative prediction between their model and tradition model is presented. They observed that their mode is giving prediction efficiency and the accuracy. [Daniel Yoas *et al.* (2014)] proposed Simple Moving Average (SMA) and Exponential Moving Average (EMA) models for prediction network traffic. They collected real network traffic from web services to test their model. Their attention was test if these models can help to forecast long-term prediction. From the experiments, they observed these models are not appropriated for long-term prediction. [Cortez *et al.* (2007)] proposed Neural Network Ensemble (NNE) ARIMA and Holt-Winter approaches with time series model for predicting Internet traffic. Authors collected data from Internet service provider based on TCP/IP protocol. They applied their techniques to data obtained at different intervals of time such as five minute, one hour a nd one day. They used MAPE metric to measure prediction performance of their techniques. A comparative prediction result between NNE, ARIMA and Holt-Winter methods is presented. They shared that the Holt-Winter approach is better to predict network at one day time scale but the ENN and ARIMA approaches achieved better prediction at five minute and one hour time interval but ARMA model is very complex than NNE approach. They noted that the NNE is more accurate overall. [Theyazn *et al.* (2016)] propose an integrated model that combines clustering with Weighted Exponential Smoothing (WES) and AutoRegressive Moving Average (ARMA) models to enhance prediction of packets loading volume in the network traffic. The data set of our research is attained from real network traffic from the WIDE backbone network.The experimental results show that the proposed model can be an effective way to improve prediction accuracy achieved with help of clustering.

3. Methodology

In this section we present and describe the methodology of our proposed model followed by algorithm.

3.1. *Proposed Model*

HTES model is one of the advance exponential smoothing models. Figure 1 demonstrates the proposed model of HTES model with clustering approach. Smoothing parameters at level and trend are selected. The MSE performance measure is used to check the convenient smoothing parameters for this data. In order to enhance the HTES model the clustering approach is applied. The centriods of cluster numbers have been selected. The prediction results that are obtained from existing HTES model are integrated with centriods clustering numbers that are obtained from clustering approach.The novelty of our proposed model is an integration of the prediction results that were obtained from HTES prediction model with centriod of clusters. The enhanced prediction is a function of existing prediction value and appropriate cluster centriod value as follows. $EP_i = f(P_i, C_i)$ where P_i is a prediction value that is obtained by conventional HTES model. C_i is the centriod of the cluster to which the i^{th} sample belongs to. It is observed that the proposed model is outperform all alternative models.

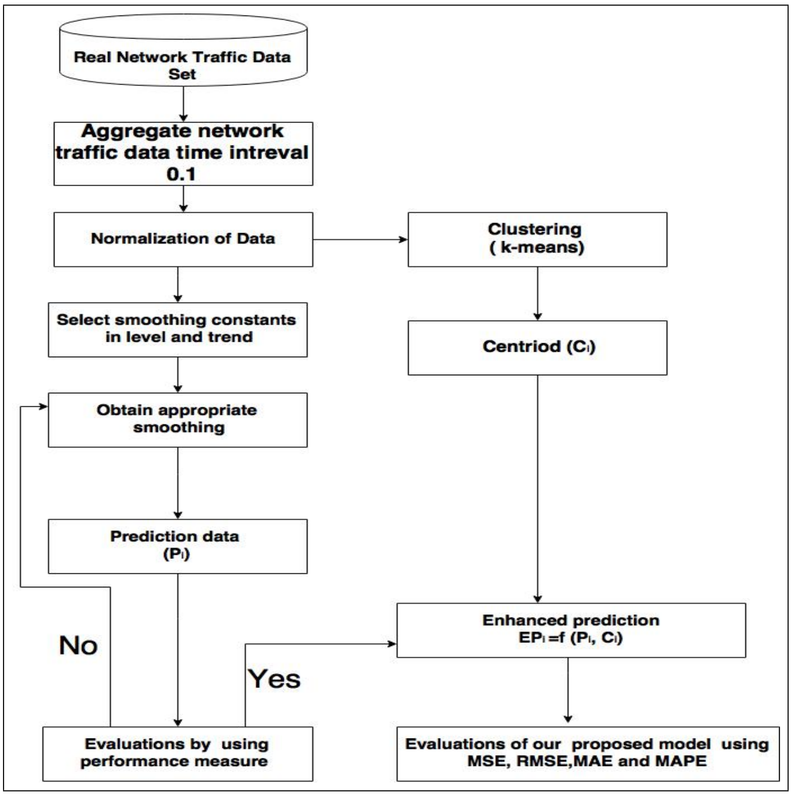

Fig. 1. Proposed model.

3.2. *Algorithm*

Let, S_i be the sample of i^{th} day, K be the number of clusters and K_i be the i^{th} cluster, C_i is the centriod of i^{th} cluster. Let P_i be the prediction for i^{th} sample obtained using HTES conventional time series prediction model and EP_i is an enhanced prediction for i^{th} sample obtained using the proposed model.

(1) Cluster network traffic flow data S_i for network traffic data using clustering techniques: k-means
(2) Apply conventional time series prediction models such as Holt-Trend model.
(3) From network traffic data obtain prediction P_i using Holt-Trend model.
(4) Modify P_i using C_j is centriod of the j^{th} cluster
$EP_i = f(P_i, C_j)$.

3.3. *Data Set*

The data set of our research is obtained from the real network that is Lawrence Berkeley Laboratory.These traces each contain an hour's worth of all wide-area traffic between Digital Equipment Corporation and the rest of the world. Detailed information about the Lawrence Berkeley Laboratory network trace is available in their homepage (http://ita.ee.lbl.gov/). In this paper, we implemented DEC-Pkt1 on time stamps have millisecond precision.

3.4. *Normalization*

The complexity and richness of network traffic has posed same obstacles in finding at the regularity explanatory principal of flow network traffic. Nevertheless, the discovery of transformation behavior in network has provided hope to enhance network traffic prediction models.We used various normalization methods such as natural logarithm and z-scores methods. It observed that the Min-max normalization is better, due to most of normalization techniques works within the range 0 and 1 but Min-max normalization method give any range of depend on data. Min-max normalization method employed for scaling our data. This method transformed data in the range of 1 to 2 scales. Figure 2 displays original data after normalization. Min-max normalization equation are defined as follows [Rob J. Hyndman *et al.* (2016)]:

$$X_n = \frac{X - X_{min}}{(X_{max} - X_{min}}(New_{max_X} - New_{min_X}) + New_{min_X} \tag{1}$$

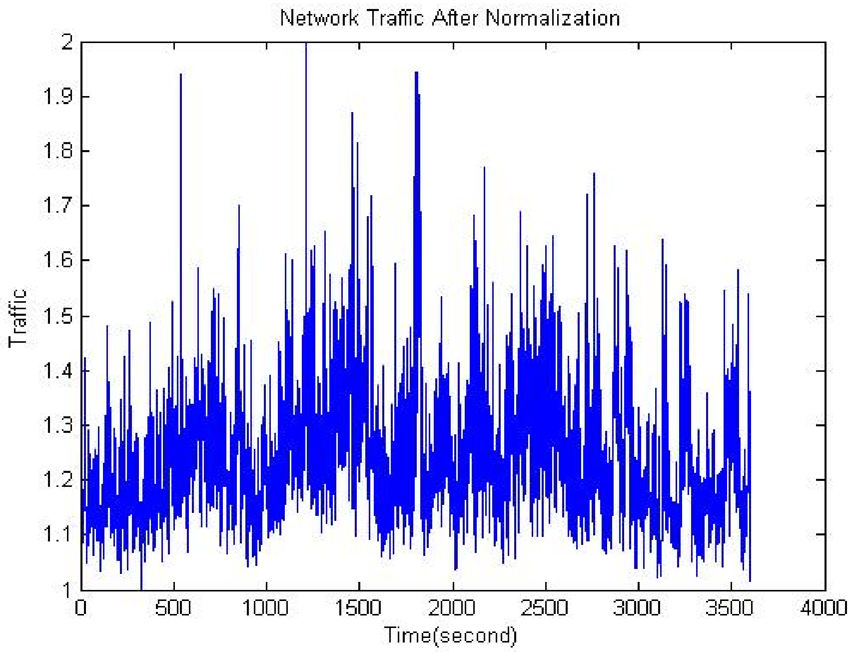

Fig. 2. Network Traffic data after Normalization

3.5. *Performance Metrics*

To evaluate the prediction model four error indicators were used. Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are applied as performance indices. These standard indicators methods are defined as follows [Rob J. Hyndman *et al.* (2016)]:

$$MSE = \frac{1}{N}\sum_{k=1}^{n}(x_t - \overline{x_t})^2 \tag{2}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{n}(x_t - \overline{x_t})^2} \tag{3}$$

$$MAE = \frac{1}{N}\sum_{k=1}^{n}(|x_t - \overline{x_t}|) \tag{4}$$

$$MAPE = \frac{\sum |(x_t - \overline{x_t})/x_t|}{n} * 100 \tag{5}$$

3.6. *Holt-Trend Exponential Smoothing Model*

Holt-Trend Exponential Smoothing (HTES) model uses a trend estimator that changes over time. Furthermore, when the parameters $\beta 0$ and $\beta 1$ are slowly changing over time. Holt-Trend Exponential Smoothing model can be employed to the time series observations when neither $\beta 0$ or $\beta 1$ is changing over time, regression can be used to forecast future values of y_t. Holt-Trend Exponential smoothing method has two smoothing constants, denoted by α and β. There are two estimators ℓ_{t-1} and b_{t-1}. The ℓ_{t-1} refer to the estimate of the level of the time series constructed in time period $t-1$ (This is typically called the level component). The b_{t-1} refers to the estimate of the growth rate of the time series constructed in time period $t1$ (This is typically called the trend component) [Rob J. Hyndman *et al.* (2016)].

Procedures of Holt-Trend model is as follows [Rob J. Hyndman *et al.* (2016)]:

(1) Initial estimates ℓ_0 and b_0 by fitting.

$$b_t = y_2 - y_1 \tag{6}$$

(2) Estimates ℓ_t and b_t by using some predetermined values of smoothing constants.
Level estimate:

$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \tag{7}$$

Trend estimate:

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta) b_{t-1} \tag{8}$$

(3) Forecasting future data by fitting this equation:

$$F_{t+m} = \ell_t + m b_t \qquad where,\ (m = 1, 2, 3, \ldots) \tag{9}$$

3.7. *K-means Clustering Approach*

The k-means clustering approach is one of the simplest unsupervised learning approaches that are used to solve the well known clustering problem. The procedures of k-means clustering is very simple and easy to classify a specified data set during a certain number of clusters (assume k clusters). Further, the k-means approach is dependent on the distances of the object from the available distributions for assigning a data object to a cluster. MacQueen is proposed Euclidean distance for measuring the distances of the object [MacQueen*et al.* (1967)]. The basic definition and approach of k-means was made efficient by. They presented the simple k-means approach which help to clusters objects in such a way that the sum of squares distance among objects within each cluster is minimized. The main objectives of k-mean is to assign n objects from x to k clusters. The k-means process starts by randomly selecting k objects as the centroids of the k clusters. Further, objects values are assigned to one of the k clusters which depend on the minimum value of the distance $d(x, c_i)$ between an object x and the centroid of cluster c_i. Frequently, the centroid of a clusters and the objects values are represented by vectors. When all the objects values are assigned to different cluster numbers new centroid vectors of the clusters are re-calculated as means of all the objects in the clusters. Furthermore, this process will be terminated when the centroids of the clusters become steady. K-means clustering equation is defined as follows [MacQueen*et al.* (1967)].

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (\|x_i - v_j\|)^2 \tag{10}$$

Where, $(\|x_i - v_j\|)$ is the euclidean distance between data point x_i and v_j. C_i is the number of the data point in J cluster and c is the number of cluster centers.

4. Experimental Setup

The TCP traffic flow data is collected form Lawrence Berkeley Laboratory. We aggregate network traffic data with time interval 0.1. The aggregation lie on sum of a regular packets in same periods for eliminating of high frequency noise. All the associated programs are written in Matlab. The detailed description of results obtained from conventional model and proposed model is presented in subsequent subsections.

4.1. *Results of Holt-Trend Exponential Smoothing (HTES) model*

The Holt-Trend model has two smoothing constants at level and trend. The parameter values from 0.1 to 0.9 for level and 0.1 to 0.30 for trend are examined. The MSE metric is used to measure these parameters for finding appropriate values of parameters of level and trend. The parameters $\alpha = 0.5$ at level and $\beta = 0.20$ at trend are selected. According to the performance measures the $\alpha = 0.5$ for level and $\beta = 0.20$ for trend outperform among other parameters. Furthermore, $\alpha = 0.5$ and $\beta = 0.20$ have less prediction error.

Table 1 displays the results that are obtained from HTES model when the parameters are $\alpha = 0.5$ and $\beta = 0.20$. The results show that the HTES model is suitable for predicting network traffic. Figure 3 illustrates graphically the prediction performance of HTES model for network traffic prediction in different time interval.From the figure, the original network data is displayed on green color, as well as the prediction for THIS model is shown blue color. The prediction errors obtained from conventional HTES model is shown in figure 4.

4.2. *Results of proposed model*

We applied k-means clustering approach to enhance HTES conventional prediction model. The k-means clustering approach is implemented on traffic data with time interval 0.1s. The Min-max normalization method is used for normalization purpose. We initiate the k- means clustering by cluster numbers and size of data. We started with 3 numbers of clusters. After discovering that there exist one or more clusters that have less objects, we decided to increase cluster numbers till we obtain all clusters with appropriate objects. Finally, we observed that clustering of network data to 5 clusters is appropriate. The centriods of appropriate clusters and predictions obtained HTES model is used to calculate enhanced predictions.

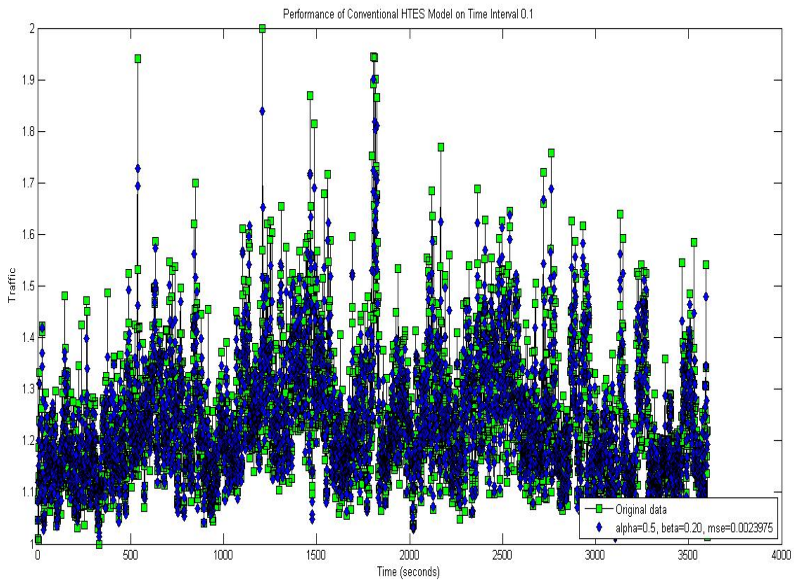

Fig. 3. Prediction performance of HTES model with time interval 0.1

Table 1 summarizes the results of proposed model against different existing models. According to different performance measures, we observed that the our proposed model is outperform over various existing prediction models. A comparative between proposed model and different existing models with time interval 0.1 is presented. Clearly, the results of proposed model seen that can effectively improve the accuracy of network traffic prediction. The generalized intelligent model performs better than the existing time series models.

We observed that the prediction errors of proposed model is 0.0010, 0.00087 0.0319, 0.0229 and 1.8107 according to MSE, NMSE, RMSE, MAE and MAPE indicators respectively on time interval 0.1. Figure 5 illustrates graphically representation of proposed model.The original network data on (green), as well as the prediction form proposed model (blue). This demonstrates that our proposed algorithm is a high accurate for improving network traffic prediction. Figure 4 displays the errors obtained from proposed model.

Table 1. Comparison of different models with our model for prediction of network traffic with time interval 0.1s

Models	MSE	NMSE	RMSE	MAE	MAPE
Flexible Neural Network(FNT) model	Not used	0.0629	Not used	Not used	Not used
Relevant Local (LSSVM+PS) model	Not used	0.0034	Not used	Not used	Not used
GP+PSO model	Not used	Not used	Not used	0.0101	4.00
Local Relevance Vector Machine+PSO model	Not used	0.0088	Not used	Not used	Not used
Holt-Trend model	0.0023	0.0022	0.0489	0.0228	2.8579
Our proposed model	**0.0010**	**0.00087**	**0.0319**	**0.0229**	**1.8107**

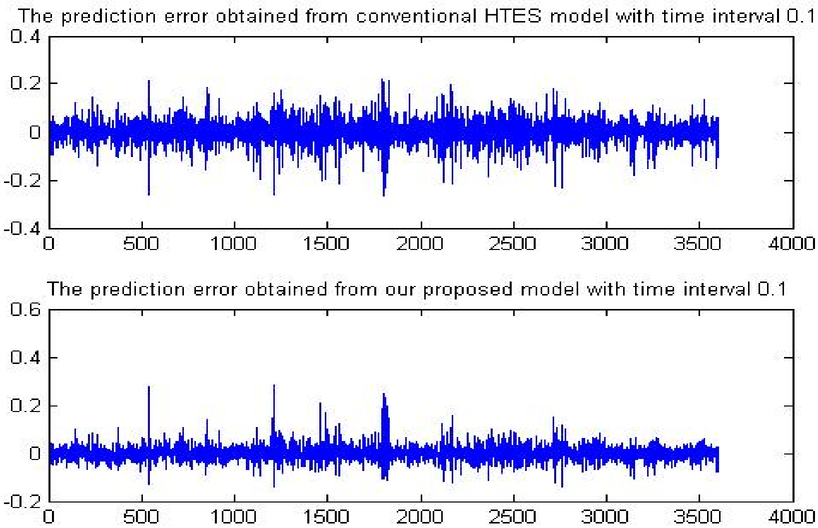

Fig. 4. The prediction errors of our model

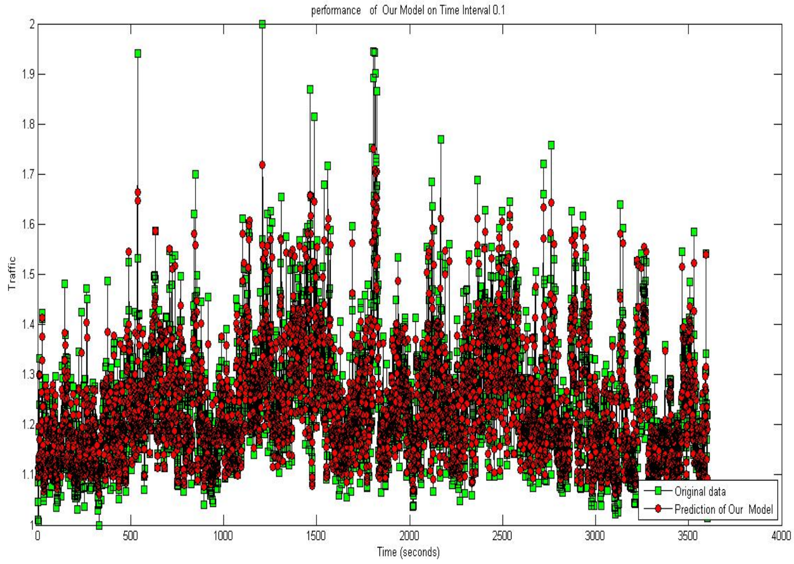

Fig. 5. Prediction performance of our model with time interval 0.1

5. Conclusion

An intelligent model is proposed to predict the network bandwidth utilization. It can assist to enhance the efficiency of the resource utilization and the scheduling of decisions of traffic data passing on high-bandwidth networks. Traffic data sharing leads to the rise in Internet for large scientific collaborations that generate huge volume of traffic data, it is challenging to network engineering for enhancing and increasing efficiency of network resources on a shared network. Furthermore, bandwidth utilization prediction has become very important for plan and designing the network. We observed that the network traffic characteristic is very complex in nature, it has more fluctuation, thus the Min-max normalization method is applied to reduce this complexity. The novelty of proposed model is the use of applied clustering approach for enhancing HTES existing model in network traffic prediction. The experimental results show that the proposed model can be successfully used for the prediction of network traffic for bandwidth utilization. A comparative prediction results between the different conventional models and the proposed model is presented. MSE, NMSE, RMSE, MAE and MAPE metrics show the prediction errors of our proposed model are 0.0010, 0.00087 0.0319, 0.0229 and 1.8107 on time interval 0.1. The graphical representation of the performance improvement shows the results of the proposed model which can be investigated. It is observed that the prediction accuracy of the propose model is superior to the various existing models. In future work we will use hadoop technology for more enhancement of network traffic prediction.

References

Bin Yang and Mingyan Jiang (2016).A Forecasting Model for Data Center Bandwidth Utilization. *SAI Intelligent Systems Conference, Conference on IEEE*, : 3039–3046.

Cisco (2008). *Trends and analysis Cisco visual networking white paper*, OTexts.

Cortez. Rio .M, Rocha .M and Sousa .P (2007).Internet Traffic Forecasting using Neural Networks. *Neural Networks, 2006. IJCNN '06. International Joint Conference on IEEE*, : 2635–2642.

Daniel Yoas, Greg Simco (2014).Using Long-Term Prediction for Web Service Network Traffic Loads. *International Conference on Information Technology Conference on IEEE*, : 21–26.

Feng Huifang, Yantai Shu (2013).Network traffic analysis and prediction based on APM. *Pervasive Computing and Applications (ICPCA), 6th International Conference on IEEE*, : 275–280.

Hartigan J.A, Wong M. A. (1979).Algorithm as A k-means clustering algorithm. *International Journal of the Royal Statistical SocietySeries*, **28**: 100-108.

Jian Kuang, Dongwei Zhai, Xinyu Wu and Yanwen Wang (2013).A Network Traffic Prediction Method Using Two-Dimensional Correlation and Single Exponential Smoothing. *Communication Technology (ICCT), International Conference on IEEE*, : 403–406.

J. Park, Raza S. M, Thorat .P and Kim D. .S (2015). Network Traffic Prediction Model Based on Training Data. *Computational Science and Its Applications, International Conference on Springer*, : 117–127.

Lun Zhang, Qiuchen Liu, Wenchen Yang, Nai Wei and Decun Dong (2013). An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction. *International Conference of Transportation Professionals, 13th COTA Conference on Elsevier*, : 654–660.

Man-Soo Han (2014). Dynamic bandwidth allocation with high utilization for xg-pon. *International Conference on Advanced Communication Technology, 16th Conference on IEEE*, : 994–997.

MacQueen, James (1967).Some methods for classification and analysis of multivariate observations. *In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, **14**.

Poo KuanHoong, Ian K.T, Tan and Chee-YikKeong (2012). BitTorrent Network Traffic Forecasting With ARIMA. *International Journal of Computer Networks and Communications (IJCNC)*, : 143–146.

Rob J. Hyndman (2016). Forecasting principles and practice. *OTexts*, .

Samar Raza Talpur, Tahar Kechadi (2016).Ensemble of Flexible Neural Tree and Ordinary Differential Equations for Small-time Scale Network Traffic Prediction. *International Journal of Computer*, **12**: 195–201.

Theyazn H.H.Aldhyani, Manish R.Joshi (2016). An Integrated Model for Prediction of Loading Packets in Network Traffic. *International Information and Communication Technology for Competitive Strategies Conference on ACM*, **2**: 654–660.

Tao Peng, Zhoujin Tang (2015).A Small Scale Forecasting Algorithm for Network Traffic based on Relevant Local Least Squares Support Vector Machine Regression Model. *International Journal of Applied Mathematics and Information Sciences*, **9**: 653–659.

Yuehui Chen, Bin Yang, Qingfang Meng (2012).Small-time scale network traffic prediction based on flexible neural tree. *International Journal of Elsevier* , **12**: 274–279.

Yuehui Meng Qing Fang, Chen Yue-Hui, Feng ZhiQuan (2013).Small-time scale network traffic prediction based on flexible neural tree. *International Journal of Acta Phys. Sin.* , **15**: 150509.