Qian GAO, Guangxia LI

Institute of Communication Engineering, PLA Univ. of Sci. & Tech.

A Novel Traffic Prediction Scheme for Broadband Satellite Communications System

***Abstract**. Multi-user traffic prediction is an important issue in broadband satellite communications system. In view of multiple UTs (User Terminal) which are treated by a single ST (Satellite Terminal), a novel NMF-based (Nonnegative Matrix Factorization-based) multi-user traffic prediction method is proposed. Compared with previous prediction schemes, the new method can reduce computational complexity and remain comparable prediction accuracy. Simulation results based on real traffic data testify the validity of the proposed method.*

***Streszczenie.** Przedstawiono metodę prognozowania przesyłu danych w satelitarnym szerokopasmowym systemie komunikacyjnym. Nowa metoda traktuje każdego z wielu użytkowników sieci jako pojedynczy terminal satelitarny. (**Nowa metoda prognozowania przepływu danych w szerokopasmowym satelitarnym systemie komunikacyjnym**)*

Keywords: Broadband satellite communications; Nonnegative matrix factorization; Traffic prediction
Słowa kluczowe: satelitarny system komunikacyjny.

1 Introduction

In recent years, the demand for broadband communication services has increased rapidly. Satellite communications network plays a vital role not only by providing broadband access directly, but also being part of the worldwide core network [1]. Broadband satellite communications system is experiencing rapid development. More and more people will choose the satellite links to access network. In the face of limited satellite bandwidth resources and users' increasing demands, it becomes a significant issue to realize reasonable and efficient bandwidth allocation among users for broadband satellite communications system. The easiest way is to assign fixed bandwidth to each user on the basis of peak rate. Because of the dynamic bandwidth demand of users, this way will lead to the waste of valuable bandwidth resources. Therefore, it becomes more and more important to study the most effective dynamic bandwidth allocation scheme so as to increase the resource utilization.

For bandwidth allocation, researchers have proposed a variety of effective schemes [2-4]. But almost all the previous schemes perform the prediction based on a specific model for each traffic trace. Then the model is trained using part of the traffic data. Finally the prediction is obtained using the trained model. For example, in the scheme proposed by Secchi R and Barsocchi [2], network traffic is modeled as linear dynamic system driven by independent random process. BP (Back Propagation) neural network is adopted to model the traffic by Jiang [3]. Delli [4] designs the bandwidth allocation mechanism on the basis of MMPP (Markov Modulated Poisson Process) model.

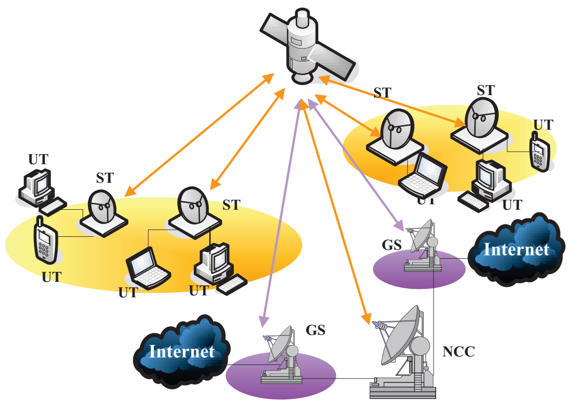

Fig.1. Broadband satellite communications system architecture

In the satellite uplink, the data of UTs are aggregated by STs. Then via onboard transponder, the data are transmitted to the GS (Gateway Station) in the ground. At the end, data reach to the ground IP backbone network. The transmission in satellite downlink is reverse. During the whole transmission, many operations are centralized processed by NCC (Network Control Center) in the ground, such as traffic control. Compared with ground network, the most significant difference lies in longer transmission delay. From ST to NCC, transmission delay of single hop is about 250ms in GEO (Geostationary Earth Orbit) broadband satellite communications system. To a large extent, one of the consequences caused by delay is that the bandwidth allocation orders provided by NCC are based on the traffic prediction of next period. This architecture also results in the demand for accurate network traffic prediction.

In fact, it is not a brand-new issue for the prediction of network traffic. There has been a variety of prediction methods proposed. These models can be roughly divided into two categories:

- Time series model. There are Markov model, ARMA (AutoRegressive Moving Average) model, ARIMA (AutoRegressive Integrated Moving Average) model, FARIMA (Fractional AutoRegressive Integrated Moving Average) model and so on [5-7].
- Prediction model based on neural networks or fuzzy neural network [8,9].

In addition to the above models, in order to eliminate the long-range dependent characteristics existing in the self-similar network traffic, the signal analysis methods based on DWT (Discrete Wavelet Transform) [9] and EEMD (Ensemble Empirical Mode Decomposition) [10] have been used for network traffic prediction. The prediction method based on ANN and adaptive template matching is also be discussed [11].

However, the methods mentioned above only consider the single-user prediction situation. In the uplink, each ST is responsible for handling multiple UTs' traffic. At this time, if every UT's traffic is treated individually, it will bring too much computational burden to the system. In addition, there is correlation between the traffic of different UTs. For example, in DVB (Digital Video Broadcasting) system, it is reasonable to infer that the majority of individual users watch video at the leisure time in evening. Therefore, this correlation can be analyzed and removed so as to compress the data before performing the prediction.

This paper focuses on the predictions of multiple UTs' traffic which are processed by a single ST in the uplink of broadband satellite communications system. Based on the research of spatial distribution characteristics of traffic data, a novel method for multiple signals prediction method using NMF (Nonnegative Matrix Factorization) is proposed.

The structure of this paper is as follows: Section 2 introduces the basic knowledge of NMF and NMF with continuity constraints briefly; Section 3 describes the proposed NMF-based multi-user traffic prediction method; Section 4 validates the accuracy of proposed method through simulations; the conclusion is given in Section 5.

2 NMF and NMF with continuity constraints

In this section, the basic knowledge of NMF is briefly reviewed in the first place. Then the NMF method with time domain continuity constraints is introduced and discussed.

2.1 Nonnegative Matrix Factorization

NMF is a linear subspace matrix factorization method proposed by Lee et al [12]. Compared with the other matrix factorization methods, such as PCA (Principal Component Analysis), SVD (Singular Value Decomposition), the most significant feature of this method is adding the nonnegative constraints to the factorization results. The factorization form is shown in equation (1) as follows:

$$\mathbf{V} \approx \mathbf{WH} \tag{1}$$

which can also be presented as:

$$v_{ij} \approx \left(\mathbf{WH}\right)_{ij} = \sum_{r=1}^{R} w_{ir} h_{rj} \tag{2}$$

where $\mathbf{V}$ is a $M \times N$ nonnegative matrix. The matrix $\mathbf{W}$ which composes of R column vectors is regarded as the basis matrix. $\mathbf{H}$ is called the encoding matrix which can be seen as the projection of $\mathbf{V}$ on $\mathbf{W}$. The nonnegative constraint can be represented as $w_{ir} \geq 0$, $h_{rj} \geq 0$. In addition, R meets the following condition:

$$\left(M + N\right) R < MN \tag{3}$$

From equation (1), we know that each column of $\mathbf{V}$ can be represented by the combination of the column vectors in $\mathbf{W}$. Furthermore, this combination is additive combination, because all the elements in $\mathbf{W}$ and $\mathbf{H}$ are positive. Due to this attribute, the NMF can be viewed as a method which is compatible with the intuitive notion of combining parts to form a whole [12]. In another words, the characteristics in $\mathbf{V}$ are carried by the column vectors of $\mathbf{W}$ with physical meanings.

As the dimension of the subspace constructed by $\mathbf{W}$ is smaller than that of $\mathbf{V}$, by projecting the matrix $\mathbf{V}$ onto the subspace formed by $\mathbf{W}$, the data can be compressed. Taking advantage of this property before prediction for multi-user, the computational complexity can be reduced.

For the specific factorization steps of NMF, two methods have been brought forward by Lee et al [13]: Euclidean distance-based error minimization method and Kullback-Leibler divergence-based error minimization method, which are presented by equation (4) and (5) respectively:

$$D_E\left(\mathbf{V} \| \mathbf{WH}\right) = \sum_{i,j} \left(v_{ij} - \left(\mathbf{WH}\right)_{ij}\right)^2 \tag{4}$$

$$D_{KL}\left(\mathbf{V} \| \mathbf{WH}\right) = \sum_{i,j} \left[v_{ij} \log \frac{v_{ij}}{\left(\mathbf{WH}\right)_{ij}} - v_{ij} + \left(\mathbf{WH}\right)_{ij} \right] \tag{5}$$

The Euclidean distance-based error minimization method is the MLE (Maximum Likelihood Estimation) of $\mathbf{W}$ and $\mathbf{H}$ in the presence of additive Gaussian noise. Moreover, the Kullback-Leibler divergence-based error minimization method is the MLE when observation data $\mathbf{V}$ are generated by the Poisson process with mean value $\left(\mathbf{WH}\right)_{ij}$. And Kullback-Leibler divergence the divergence is more sensitive to low-energy observations. In traffic prediction, it is considered that low-frequency part carries most of traffic trend information. Relative to the high-frequency part, the prediction accuracy is more vulnerable to low frequency part. Thus, Kullback-Leibler divergence error minimization-based NMF method is utilized in prediction method proposed in this paper.

2.2 NMF with continuity constrains

In practice if the rows of $\mathbf{V}$ are time series, the rows of encoding matrix $\mathbf{H}$ can be viewed as representing the time domain evolution of $\mathbf{V}$ on the subspace of $\mathbf{W}$. In many applications, such as speech signal spectrum analysis, time domain continuity is an inherent property of the data which naturally leads to time domain continuity in the encoding matrix. However the time dimension continuity of encoding matrix hasn't been taken into consideration in the original NMF method. In the original NMF, every column of $\mathbf{V}$ is assumed as individual observation. To cope with this problem, Tuomas et al has proposed a NMF method which incorporating the continuity constraint [14].

In this scheme, temporal continuity is favored by using a cost term which is the sum of squared differences between the gains in adjacent columns, as shown in equation (6):

$$c\left(\mathbf{W}, \mathbf{H}\right) = c_r\left(\mathbf{W}, \mathbf{H}\right) + \alpha c_t\left(\mathbf{H}\right) \tag{6}$$

where $c_r\left(\mathbf{W}, \mathbf{H}\right)$ is the reconstruction error item as shown in equation (5). $c_t\left(\mathbf{H}\right)$ is the time domain continuity constraint item on encoding matrix denoted by equation (7). α is the weight factor.

$$c_t\left(\mathbf{H}\right) = \sum_{r=1}^{R} \frac{1}{\sigma_r^2} \sum_{j=2}^{N} \left(h_{r,j} - h_{r,j-1}\right)^2 \tag{7}$$

where σ_r is the standard deviation of the r th row in encoding matrix.

By adding time domain continuity constraints items, the encoding matrix obtained from NMF is presented with better continuity in time domain. When applied for prediction, it is no doubt that this continuity will improve the prediction accuracy. In next section, we will propose a new NMF-based scheme with time domain continuity constraint for multiple users traffic prediction.

3 NMF-based traffic prediction

The proposed method is used to predict network traffic of multiple users. The flow chart of method is shown in Fig.2. The prediction process can be divided into two phases, which are the training phase and the prediction phase.

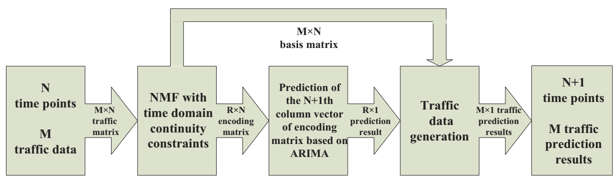

Fig.2. Flow chart of proposed traffic prediction method

3.1 Training phase

1) Assume that there are M users whose traffic data need to be predicted. First of all, the $M \times N$ traffic matrix is constructed using N former observations of the M users;

2) Employ NMF with time domain continuity constraints to analyze the traffic matrix, resulting in the $M \times R$ basis matrix and $R \times N$ encoding matrix;

3) Adopt ARIMA model to modeling each row vector of encoding matrix.

3.2 Prediction phase

Given the former k values of each row in the encoding matrix have been obtained, then:

1) Predict the values of encoding matrix at time point $k+1$ for each row based on the trained ARIMA models and the previous values of the encoding matrix. The prediction results are denoted as $\hat{\mathbf{h}}_{k+1}$;

2) Multiply the prediction results with the basis matrix which results in the prediction results of the traffic data for the M users at time point $k+1$:

$$\hat{\mathbf{v}}_{k+1} = \mathbf{W}\hat{\mathbf{h}}_{k+1} \tag{8}$$

where $\mathbf{v}_{k+1}$ is the traffic prediction results at $k+1$.

3) Update the encoding matrix when the traffic data at time $k+1$ has arrived by the following equation:

$$\mathbf{h}_{k+1} = \frac{\mathbf{v}_{k+1}}{\mathbf{W}} \tag{9}$$

where $\mathbf{v}_{k+1}$ are the newly arrived traffic data for the M users at $k+1$.

4) Return to step 1) to carry out traffic prediction for the next time point.

In the proposed method, ARIMA model is utilized for the prediction of the encoding matrix. The reasons for choosing this model due to that: compared with ARMA model, ARIMA model can model the nonstationary characteristics of signals better. In addition it needs less computation to perform prediction than FARIMA model. Finally, as a mature prediction model, the prediction accuracy can be ensured.

4 Experiments and analysis

The experiments are carried out based on one segment of Abilene III network traffic data which is taken from the router node in Abilene network on June 1st, 2004. The trace statistics summary of Abilene III is shown in Table 1.

Table 1. Trace statistics summary of Abilene III

Bandwidth	Total packets	Total flows	Utilization	Flows in progress	MTU
10Gbps	156M	683K	19%	62000	9000

Note: MTU is the Maximum Transmission Unit.

Preprocessing of the original data is performed in the first place. More specifically, 10 channels are extracted from the trace. And each channel contains the observation of 3000s with the observation interval of 5s. So there are 600 observation points for each channel. After that the obtained traffic data are normalized. In experiments, the former 500 traffic data are used as the training data. And the remaining 100 traffic data are left for prediction test. The 10 channels traffic are presented in Fig.3.

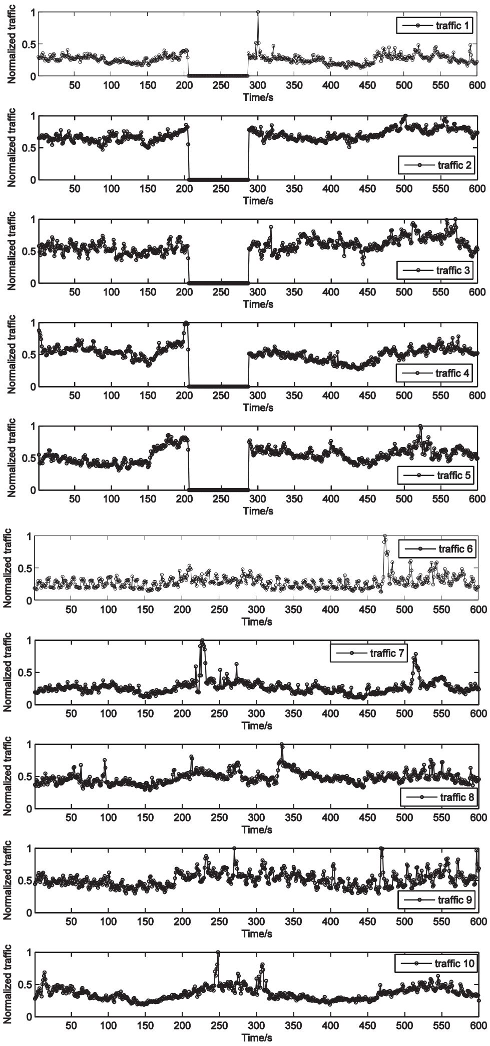

Fig.3. The 10 channels traffic data taken from Abilene III

4.1 The training and prediction process

In the training phase, the former 500 traffic data of each channel are extracted to form the 10×500 traffic matrix $\mathbf{V}$. Then the $\mathbf{V}$ is processed by time continuity constrained NMF resulting in the basis matrix and encoding matrix. The dimension of the basis matrix is set to 6. So the size of the obtained basis matrix is 10×6 and that of the encoding matrix is 6×500. The weight factor α adopts empirical value 1.2.

After the encoding matrix has been achieved, the 6 row vectors of encoding matrix are modeled by the ARIMA models respectively. The ARIMA model contains 3 parameters which can be represented as ARIMA(p,d,q). By repetitious experiments of different combinations of these parameters, the optimal parameters are chosen depending on the data fitting error of model. The searching range of each parameter is:

(1) The parameter candidates for p and q are: 1, 2, 3;

(2) The parameter candidates for d are: 0, 1.

Finally, the selected parameters are shown in Table 2.

Table 2. ARIMA model parameters selection results for 6 rows of encoding matrix

Channel number	1	2	3	4	5	6
Model parameters	(2,0,3)	(2,1,3)	(2,0,3)	(1,0,2)	(2,1,3)	(1,0,3)

Then encoding matrix for the testing data is obtained by the following equation:

(10) $$\mathbf{H}_t = \frac{\mathbf{V}_t}{\mathbf{W}}$$

Through the trained ARIMA models, the prediction results of the testing data are calculated and represented by $\mathbf{H}_t'$. And the prediction results are shown in Fig.4.

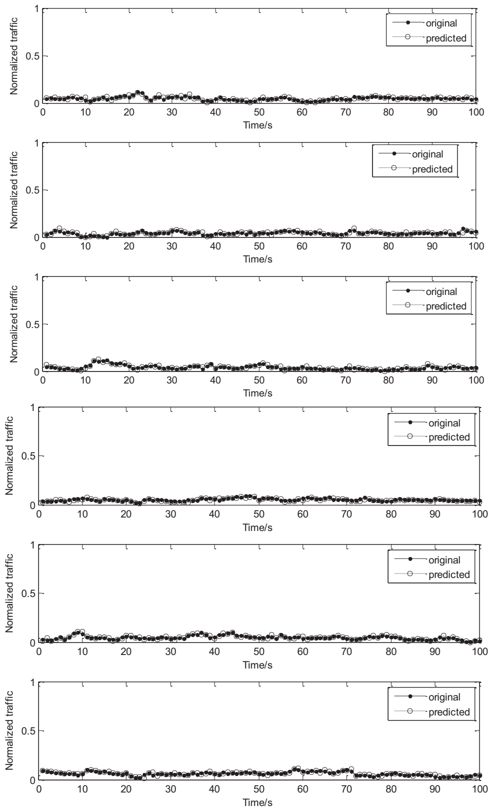

Fig.4. The prediction results for each row of encode matrix based on ARIMA

As the continuity constrained is included in the encoding matrix, the variance in each row of $\mathbf{H}_t'$ is smaller than that of the original traffic which can be seen for Fig. 4 and Fig. 3. Then it will be much easier to be predicted by the ARIMA model.

In virtue of the predicted testing encoding matrix $\mathbf{H}_t'$ and basis matrix $\mathbf{W}$, the prediction results for each channel of the original traffic data are obtained.

4.2 The analysis the prediction results

To measure the prediction performance, we choose MSE (Mean Squared Error) as the evaluation criterion. Typically, the definition of the MSE is:

(11) $$e = \frac{1}{K}\sum_{i=1}^{10}\sum_{t=1}^{K}\left(\hat{v}_t^i - v_t^i\right)^2$$

where e is the MSE value of predictions, v_t^i is the actual value for the k th predicted result in i th channel and $\hat{v}_t^i$ is the corresponding predicted value. K is the number of predictions.

In order to investigate the validity of NMF-based prediction method, the ARIMA model is used directly for predicting the traffic data as comparison test. The results are shown in Table 3.

Table 3. Comparison of prediction error for different methods

Prediction method	e
NMF-based prediction	0.0651
ARIMA	0.069

As can be known from Table 3, the prediction error of NMF based method is a little bit lower than that of ARIMA based method which means the proposed method can acquire comparable prediction result as that of the ARIMA based method.

However compared with the training time of ARIMA model, the decomposition of continuity constrained NMF is a very time-saving process. As the parameter selection of ARIMA is carried out based on the results of repetitious experiments, a large amount of computation has been brought by this process. According to subsection 4.1, 18 repetitious experiments have to be performed for the parameter selection of each ARIMA model. In Table 4, the consumed time for the training of the six ARIMA models used for the encoding matrix and time used for performing the continuity constrained NMF are shown. The time is recorded based on the computer of the following configurations: Intel E7500Dual core 2.93GHz CPU, 2GB memory. And the experiments are performed based on matlab 2010a.

Table 4. Time consuming of the training process

Method	Time (s)
ARIMA for row 1	4.18
ARIMA for row 2	3.59
ARIMA for row 3	4.41
ARIMA for row 4	3.96
ARIMA for row 5	4.22
ARIMA for row 6	4.51
continuity constrained NMF	0.91

According to the experimental results, the decomposition time of continuity constrained NMF is very short compared with the ARIMA model training.

In addition, the total time used by the direct ARIMA based method for the perdition is 45.12s, while that for our proposed method is just 26.54. That means 41.18% of the computational complexity can be saved.

5 Conclusion

A novel traffic prediction scheme based on time continuity constrained NMF has been proposed in this paper. The new scheme can be used for performing the multi-channel traffic prediction in the broadband satellite communications system. Through NMF, the original multi-channel traffic data are projected onto a low dimension subspace. The data need to be predicted are reduced and prediction efficiency is improved. The experimental results show that the NMF-based prediction method presents lower computing complexity than direct traffic prediction method. At the meantime, the prediction accuracies of the two methods are comparable. These demonstrate the validity of the proposed method.

Acknowledgement

The authors wish to acknowledge the National Natural Science Foundation of China (60972061 and 61032004), National High Technology Research and Development Program of China ("863" Program) (2008AA12A204, 2008AA12Z307).

REFERENCES

[1] Markos P. Anastasopoulos, Dionysia K. Petraki, et al., Call Admission Control Scheme for Multiclass Services under Rain Fading for Satellite Networks, IEEE Transaction on Wireless communications, 8 (2009), No. 5, 2473-2483

[2] Secchi R, Barsocchi P, Davoli F, Linear quadratic control of service rate allocation in a satellite network, Communications, IET, 4 (2010), No. 13, 1580-1593
[3] Yueqiu J, Hang L, Study on Bandwidth Allocation of Broadband Multimedia Communication Satellite System, 3rd International Conference on Intelligent Networks and Intelligent Systems, (2010) ,506-508
[4] Delli Priscoli F, Pompili D, A demand-assignment algorithm based on a Markov modulated chain prediction model for satellite bandwidth allocation, Wirel. Netw., 15 (2009), 999-1012
[5] Johnson E L, Sivalingam K M, Mishra M, Scheduling in Optical WDM Networks Using Hidden Markov Chain Based Traffic Prediction, Photonic Network Communications, 3 (2001), 269-283
[6] Kirchg ssner G W J, Introduction to modern time series analysis, Springer, (2008)
[7] Yantai S, Zhigang J, Lianfang Z, et al., Traffic prediction using FARIMA models, IEEE International Conference on Communications, (1999), 891-895
[8] Satsri S, Ardhan S, Chutchavong V, et al., ANN based NGN IP traffic prediction in Thailand, International Conference on Control, Automation and Systems, (2007), 2154-2157
[9] LI R, CHEN J, LIU Y, et al., WPANFIS: combine fuzzy neural network with multiresolution for network traffic prediction, The Journal of China Universities of Posts and Telecommunications, 17 (2010), No. 4, 88-93
[10]Gao B, Zhang Q, Liang Y, et al., Predicting self-similar networking traffic based on EMD and ARMA, Journal on Communications, 32 (2011), No. 04, 47-56 (in Chinese)
[11]GAO Qian, LI Guang-xia, A Traffic Prediction Method based on ANN and Adaptive Template Matching. This paper has been submitted to *Przegląd Elektrotechniczny* with paper ID: PE2257
[12]Lee D D, Seung H S, Learning the parts of objects by non-negative matrix factorization, Nature, 401 (1999), No. 6755, 788-791
[13]Daniel D. Lee, H. Sebastian Seung, Algorithms for Non-negative Matrix Factorization, Advances in Neural Information Processing Systems, Proceedings of the 2000 Conference. 556–562
[14]Tuomas V, Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria, IEEE Transactions on Audio, Speech, and Language Processing, 15 (2007), No. 3, 1066-1074

***Authors**: Qian GAO, dr, Military Satellite Communications LAB, Institute of Communications Engineering, PLA University of Science and Technology, Biaoyin 2, Yudao Street, Nanjing, China, 210007, Email: gaoxiongmao1234@163.com;*
Guangxia LI, prof., Military Satellite Communications LAB, Institute of Communications Engineering, Biaoyin 2, Yudao Street, Nanjing, China, 210007.