Received: 7 February 2023 | Revised: 11 February 2024 | Accepted: 21 August 2024

DOI: 10.1049/cmu2.12863

IET Communications

ORIGINAL RESEARCH

Self-similar traffic prediction for LEO satellite networks based on LSTM

Yan Zhang[1] | Yong Wang[1] | Haotong Cao[2] | Yihua Hu[1] | Zhi Lin[1] | Kang An[3] | Dong Li[4]

[1]College of Electronic Engineering, National University of Defense Technology, Hefei, China

[2]Department of Computing, Hong Kong Polytechnic University, Hong Kong, China

[3]Sixty-Third Research Institute, National University of Defense Technology, Nanjing, China

[4]School of Computer Science and Engineering, Macau University of Science and Technology, Macau, China

Correspondence
Yihua Hu, Huangshan Road, Shushan District, Hefei City, Anhui Province, China.
Email: skl_hyh@163.com

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 62201592, 61901502, 61671454; Research Plan Project of NUDT, Grant/Award Number: ZK21-33; Macau Young Scholars Program, Grant/Award Number: AM2022011; Young Elite Scientist Sponsorship Program of CAST, Grant/Award Number: 2021-JCJQ-QT-048; Postgraduate Scientific Research Innovation Project of Hunan Province, Grant/Award Number: CX20220007

Abstract
Traffic prediction serves as a critical foundation for traffic balancing and resource management in Low Earth Orbit (LEO) satellite networks, ultimately enhancing the efficiency of data transmission. The self-similarity of traffic sequences stands as a key indicator for accurate traffic prediction. In this article, the self-similarity of satellite traffic data was first analyzed, followed by the construction of a satellite traffic prediction model based on an improved Long Short-Term Memory (LSTM). An early stopping mechanism was incorporated to prevent overfitting during the model training process. Subsequently, the Diebold-Mariano (DM) test method was applied to assess the significance of the prediction effect between the proposed model and the comparison model. The experimental results demonstrated that the improved LSTM satellite traffic prediction model achieved the best prediction performance, with Root Mean Squared Error values of 18.351 and 8.828 on the two traffic datasets, respectively. Furthermore, a significant difference was observed in the DM test compared to the other models, providing a solid basis for subsequent satellite traffic planning.

1 | INTRODUCTION

In July 2022, the American Satellite Industry Association (SIA) released its 25th annual report detailing the status of the satellite industry. According to the data, the total revenue of the global aerospace industry in 2021 was 386 billion USD, an increase of 15 billion USD compared to 2020, representing a growth rate of 4%. The satellite industry contributed $279 billion, accounting for 72% of the global space industry revenue. LEO Satellite Networks (LSNs) have garnered significant attention due to their advantages of low latency and wide coverage. Economic powers have proposed low orbit satellite constellation plans, joining the satellite network “space race” ambitions, with OneWeb, Kupier, and Starlink being the most rapidly developing LEO satellite constellations [1, 2]. The Starlink constellation by the American space exploration company SpaceX in particular offers users communication rates that have reached the level of terrestrial fiber optic networks.

Traffic prediction is a crucial aspect of traffic balancing and resource management in satellite networks. By forecasting the traffic demand for the upcoming time period for each satellite, the network can proactively plan and prioritize traffic transmission, actively avoiding congestion and enhancing data transmission efficiency [3–7]. However, LSNs are highly dynamic, complex, and uncertain, posing significant challenges to accurate traffic prediction. As the volume of data traffic

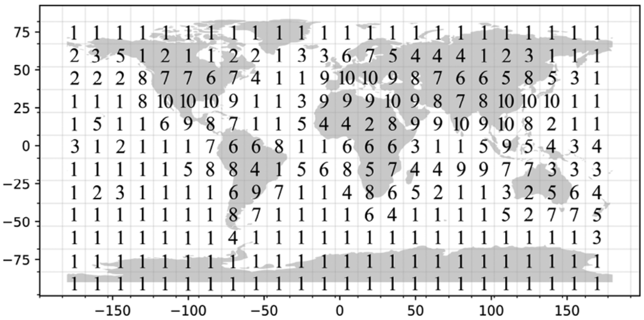


FIGURE 1 Traffic distribution density on the earth's surface (The numbers from 1 to 10 as traffic index to represent the TDD; the larger the number, the greater the TDD).

carried by LSNs continues to grow rapidly, accurate prediction of satellite traffic has gained significant attention to improve the utilization of LEO satellite bandwidth resources [8, 9].

Fortunately, the LEO satellite constellation also functions as a regular network with a high level of periodicity, correlation, and symmetry. This characteristic enables the prediction of traffic patterns for each satellite in the future, based on its specific attributes [10, 11]. Consequently, the overall efficiency of the satellite network can be enhanced [12]. The distribution of traffic within LSNs is closely tied to the geographic coverage provided by the satellites and the daily cycle variations. For instance, satellites positioned over oceans generally experience lighter loads, while those above bustling cities encounter heavier usage. Moreover, daytime traffic loads in different regions tend to be more pronounced compared to nighttime loads, owing to the influence of daily time patterns. As illustrated in Figure 1, densely populated and economically developed areas indicate a more robust terrestrial telecommunications infrastructure.

However, despite our understanding of the components and characteristics of satellite traffic, predicting the traffic of LSNs still poses two challenges. Firstly, traditional traffic prediction models like Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) can enhance prediction accuracy in terrestrial networks but are computationally and storage-intensive for satellite networks due to limited resources [13]. Secondly, when terrestrial network traffic is combined with satellite traffic, its inherent properties (self-similarity, long correlation, nonlinearity) become more complex. Therefore, it is crucial to develop a prediction model based on the self-similar characteristics of LEO satellite network traffic to advance research on satellite resource scheduling and load balancing [14]. In this study, we first analyze the properties of satellite network traffic and construct improved Long Short-term Memory (LSTM), Gated Recurrent Unit (GRU), and Sequence-to-Sequence (Seq2Seq) traffic prediction models for LSNs using neural networks. We then perform comparative experiments on two network datasets to establish a prediction method suited for self-similar traffic characteristics, laying the groundwork for subsequent traffic planning.

The remaining sections of the paper are structured as follows. Section 2 presents an overview of the related work. In Section 3, we delve into the satellite network traffic and its properties. Subsequently, Section 4 provides a detailed description of the methods employed. Section 5 showcases the analysis results and initiates a discussion on the findings. Finally, in the concluding section, we present our conclusions.

2 | RELATED WORK

With the rapid growth of data traffic carried by LSNs, accurate prediction of satellite traffic load has become crucial for guiding satellite data transmission and improving the utilization of satellite bandwidth resources. Consequently, researchers have focused on satellite network traffic prediction.

2.1 | Traffic prediction based on traditional models

Traditional models like ARMA and ARIMA [15] are commonly used for network traffic prediction. However, these models perform poorly when applied to self-similar characteristic traffic in satellite networks due to their reliance on short time series, which cannot capture the self-similarity and long memory of satellite network traffic [16].

The Fractional Autoregressive Integrated Moving Average (FARIMA) model, an improvement over the ARIMA model, has gained popularity among researchers for its ability to characterize self-similarity and long correlation [17]. However, the FARIMA model is linear in nature and fails to accurately represent the nonlinear compositional relationships in satellite network traffic. Additionally, the model complexity is high.

2.2 | Traffic prediction based on feature decomposition

An alternative approach to improve prediction involves decomposing the original satellite traffic into multiple short-term correlated data components. Gao et al. [18] employed Empirical Mode Decomposition (EMD) to decompose self-similar satellite network traffic into multiple short-term correlation components. The EMD-ARMA composite forecasting model was proposed, which predicts the decomposed components of network traffic. However, the decomposed components exhibit nonlinear characteristics, resulting in significant prediction errors.

Bie et al. [19] proposed a combined prediction model for satellite network traffic based on EMD, and regarding the decomposed high-frequency application the Fruit Fly Optimization Algorithm and Extreme Learning Machine (FOA-ELM) prediction model, for the low-frequency application ELM prediction model. However, the self-similarity of satellite network traffic is not discussed, and the model has not been validated in complex satellite network topologies.

Similarly, the wavelet transform method can be used to decompose non-smooth, long-correlated traffic into multiple

TABLE 1 Comparison of different satellite traffic prediction methods.

Categorization	Model/reference	Advantage	Disadvantage
Traditional models	ARMA, ARIMA [15]	The model construction is straightforward and exhibits strong interpretability.	Can only simulate short relevant sequences with low accuracy.
	FARIMA [17]	Able to simultaneously describe short-term and long-term processes.	The model structure is complex and cannot handle non-linear network traffic sequences.
Feature decomposition	EMD-ARMA [18]	Reduced model complexity.	The decomposed components exhibit nonlinear characteristics, leading to significant prediction errors.
	Wavelet transform [20]	Fast prediction capabilities.	Susceptible to time delay issues.
Artificial intelligence	SVR [21]	Strong generalization ability, avoiding excessive learning issues.	Different kernel functions have significant deviations and require a large amount of data.
	PCA-GRNN [16]	Taking into account the temporal and spatial correlation of satellite data, the nonlinear generalization capability is strong.	Only for short-term forecasting and high model complexity.
	GRU- Transfer learning [23]	Solved the problem of insufficient satellite traffic data.	Lack of description of satellite network traffic characteristics (self-similarity and long memory).

scales, providing effective local analysis in both the time and frequency domains simultaneously. Tian [20] employed the wavelet transform technique to decompose and reconstruct the original network traffic, extracting approximate and detailed components for fusion with a multi-model approach for network traffic prediction.

2.3 | Traffic prediction based on artificial intelligence

The advantage of traditional models in traffic prediction lies in the interpretability of the mathematics and the ability to make short-term predictions of traffic sequences. However, these models often have high computational complexity due to a large number of parameters. The emergence of artificial intelligence can address the limitations of traditional models and achieve higher flexibility and generalization in prediction, making it the current mainstream direction for satellite network traffic prediction.

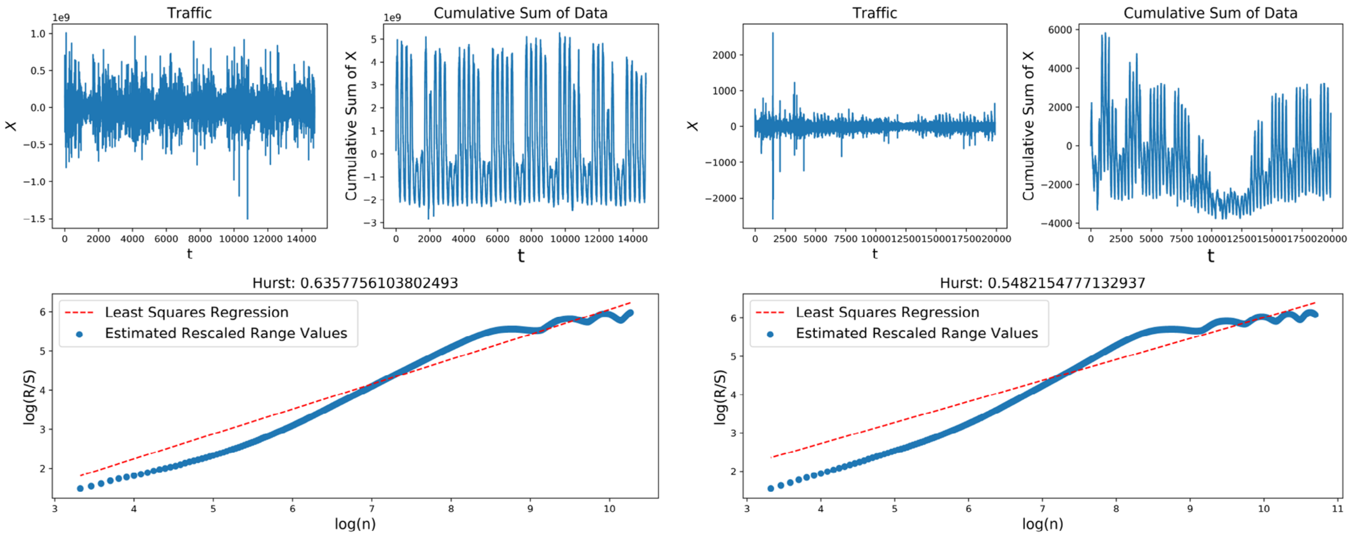
Liang et al. [21] proposed a traffic forecasting algorithm based on ant colony optimization and Support Vector Regression (SVR). While this algorithm has fast convergence and certain advantages, it requires a large amount of data. Na et al. [22] considered the traffic distribution density on the Earth's surface and introduced a distributed routing strategy based on ELM. They used ELM to predict satellite node traffic and incorporated the concept of mobile agents to determine routing information. Simulation results showed that the ELM distributed routing strategy can reduce packet loss rate and switching delay.

Deep learning models possess excellent nonlinear fitting capabilities for complex problems. Some researchers have started using deep learning models to predict satellite traffic. Liu et al. [16] designed a satellite network traffic prediction algorithm based on Principal Component Analysis and Generalized Regression Neural Networks (PCA-GRNN), considering the temporal periodicity and spatio-temporal correlation of satellite traffic. It achieves high accuracy and stability but is focused on short-term traffic prediction. Li et al. [23] proposed GRU neural network traffic prediction algorithm, which addresses the issue of insufficient online traffic data by leveraging transfer learning methods and achieves accurate satellite network traffic prediction. However, the algorithm has high computational complexity, leading to increased computational load on the satellite. Additionally, although satellite network traffic shares basic characteristics with terrestrial network traffic, it is important to consider that satellite network traffic exhibits greater time variability and complexity. In Table 1, we summarize the advantages and disadvantages of various satellite network traffic prediction methods.

3 | SATELLITE NETWORK TRAFFIC AND ITS PROPERTIES

Traffic prediction is essential for traffic load balancing and resource management in LEO satellite networks. By forecasting the traffic for each satellite in the upcoming time slot, the network can proactively compute routing tables and plan traffic transmission paths, thereby enhancing data transmission efficiency. For satellite QoS applications that pre-allocate bandwidth resources, reserving based on peak values leads to resource wastage. In contrast, dynamic allocation according to traffic prediction results improves bandwidth utilization while ensuring satellite QoS [16].

The traffic load of a satellite results from aggregating the traffic demands of all users within its coverage area. In LEO satellite networks, each satellite offers data transmission services to all user terminals within its coverage. Consequently, inter-satellite traffic demand primarily relies on traffic demands between users in their coverage areas. As the user traffic intensity within satellite coverage areas fluctuates throughout the day, inter-satellite traffic demand also varies over time.



(a) Hurst exponent of ec_data. (b) Hurst exponent of uk_data.

FIGURE 2 Hurst exponent of two datasets.

Moreover, since the coverage area of each satellite changes periodically and user traffic demand intensity between regions exhibits distinct daily cyclic characteristics, the traffic demand between satellites similarly demonstrates specific periodic fluctuations. It is reasonable to assume that satellite networks carry comparable user traffic demand to terrestrial networks. At a network's dynamic level, the fundamental characteristics of terrestrial network traffic (e.g. self-similarity, long-range dependence, burstiness) persist upon entering satellite networks through terrestrial gateways. This indicates that the fundamental characteristics of satellite network traffic remain consistent with terrestrial network traffic, albeit with heightened time-varying complexity [24]. Therefore, established terrestrial network traffic prediction methods offer valuable insights when performing satellite network traffic prediction.

The self-similarity of the traffic series is an important indicator for traffic prediction. The degree of autocorrelation of a flow series is determined by considering the correlation between data at an interval (close to 1 is positive correlation, close to -1 is negative correlation, and close to 0 indicates very low correlation). The equation for the self-similarity function is as follows:

$$\rho(h) = \frac{E\left[(X_t - \mu)\left(X_{t+h} - \mu\right)\right]}{\sigma^2}, \tag{1}$$

where X_t is the sampling point, h is the interval, μ and σ are the mean and standard deviation of X_t.

The Hurst exponent is a key metric for measuring long-term memory in traffic time series. Originally developed in hydrology, it has since been applied to various time series analyses. Specifically, a Hurst exponent greater than 0.5 indicates positive memory in the traffic series. If the series has been trending upwards or downwards in the recent past, it is likely to continue this trend in the future. The closer the Hurst exponent is to 1, the stronger the long-term memory trend. Conversely, a Hurst exponent less than 0.5 indicates reverse memory in the traffic series. A value of 0.5 signifies no memory between the series.

In the context of a satellite network, when nodes are in motion, different nodes may follow similar terrestrial trajectories. Additionally, global user traffic exhibits similar diurnal patterns influenced by human activity. Hence, satellite networks primarily carry similar or proportional user traffic demands compared to terrestrial networks. Consequently, this paper focuses on predicting traffic transmitted from terrestrial sources to satellites. To validate the presence of self-similarity and long-term memory in the selected datasets, we utilize internet traffic data from a private ISP with centers in 11 European cities (referred to as ec_data) [25], as well as internet traffic data from a UK academic network backbone (referred to as uk_data) [26]. The ec_data corresponds to a transatlantic link and was collected from 7 June 2005, 06:57 hours to 31 July 2005, 11:17 hours. The uk_data was collected between 19 November 2004, 09:30 hours, and 27 January 2005, 11:11 hours, both at 5 min intervals. The Hurst exponent values for these two datasets are depicted in Figure 2. The X represents the difference of sub-time series, t represents the amount of data, $\log(R/S)$ denotes the logarithm of the extreme range of rescaling of the series, $\log(n)$ denotes the logarithm of the length n of sub-time series, and its slope is Hurst index. The statistical analysis of the two datasets is presented in Table 2.

4 | METHODOLOGY

4.1 | Improved long short-term memory

Recurrent Neural Networks (RNNs) have found widespread applications in machine translation, speech recognition, and other fields. They are characterized by the presence of cyclic connections in the hidden layer, which allow them to utilize historical information to influence output results. As a result,

TABLE 2 Description of the two Internet traffic datasets.

Dataset	Time scale	Train length (selected)	Valid length (selected)	Test length (selected)	Total length
Ec_data	5 min	1500	690	690	14772
	1 h	125	58	58	1231
Uk_data	5 min	1500	690	690	19888
	1 h	125	58	58	1657

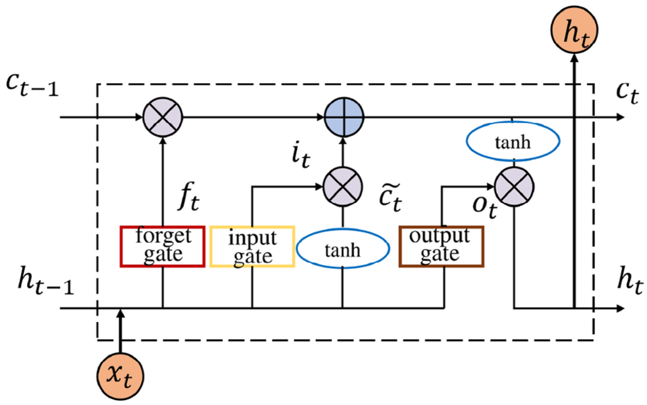


FIGURE 3 LSTM structure.

RNNs outperform traditional neural networks in processing time series data.

Long Short-Term Memory (LSTM) is an improved variant of the recurrent neural network that excels at learning long-term and short-term dependencies in time series data for processing and prediction tasks. LSTM incorporates a unique mechanism for managing thresholds [27, 28]. A typical LSTM network consists of three gates: the input gate, the forget gate, and the output gate. The input gate allows the network to process and incorporate new data from external sources. The forget gate controls the extent to which the cell state of the network unit is forgotten, thereby preserving the most relevant memory cell states. Finally, the output gate selectively outputs the computed results from the LSTM unit. The overall structure of LSTM is illustrated in Figure 3.

According to the figure, the attenuation coefficient equation of the forget gate is:

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right), \tag{2}$$

where σ is the Sigmoid excitation function, W_f is the weight matrix of the forget gate, h_{t-1} is the neuron state of the forget gate at the last moment, x_t is the input value at the current moment, and b_f is the forget bias term. The equation for the input gate is:

$$i_t = \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right), \tag{3}$$

$$\tilde{c}_t = \tanh\left(W_c\left[h_{t-1}, x_t\right] + b_c\right), \tag{4}$$

$$c_t = i_t \tilde{c}_t + f_t c_{t-1}. \tag{5}$$

First, a layer of sigmoid excitation function is passed, so that the input value i_t will be updated, and b_i is an input bias term. Then, a new candidate $\tilde{c}_t$ is selected by tanh excitation function, and the state of the old neuron is calculated and updated, and the state of the neuron c_{t-1} at the last moment is updated to c_t. Therefore, the input gate can prevent useless information from entering memory. The equation for the output gate is:

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right), \tag{6}$$

where b_o is the output gate bias term, and the output information o_t is screened out by a layer of sigmoid excitation function.

Finally, the output of LSTM is jointly determined by the output gate and the previously calculated cell state c_t. The calculation equation of h_t is as follows:

$$h_t = o_t \tanh\left(c_t\right). \tag{7}$$

When working with sequential data, incorporating the attention mechanism can automatically capture and learn sequential contextual information, thereby enhancing the accuracy of model predictions. Hence, for satellite traffic prediction, we propose an improved LSTM algorithm that integrates the attention mechanism. This mechanism is added after the LSTM layer to establish long-range dependencies among input traffic sequences and dynamically generate weights for different connections. The structure of the improved LSTM satellite traffic prediction model is illustrated in Figure 4. The learning process involves obtaining a series of attention coefficients $\{a_1, a_2, \ldots, a_t\}$ that represent the importance of each intermediate state in the output vectors of the LSTM module $\{h_1, h_2, \ldots, h_t\}$. Ultimately, the output sequence $\hat{y}$ is derived as the weighted sum of each intermediate state.

4.2 | Gated recurrent unit

Compared with LSTM, GRU has a simpler structure and fewer parameters, which can not only simulate the problem with multiple input variables, but also effectively reduce the complexity of the network [29, 30]. The structure of GRU is shown in Figure 5.

Comparing Figures 3 and 5, we can see that the LSTM has three gating units and a more complex network structure. The network structure of the GRU consists of only two gating units: the reset gate and the update gate, which correspond to r_t and u_t in Figure 4. The reset gate is responsible for controlling the

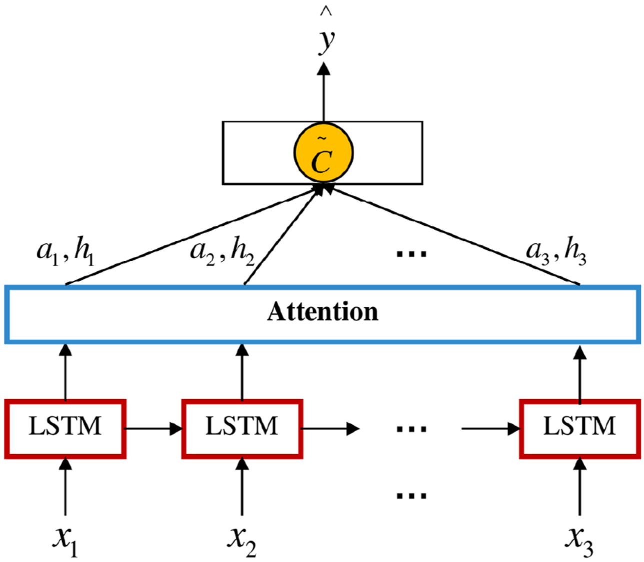


FIGURE 4 Improved LSTM structure.

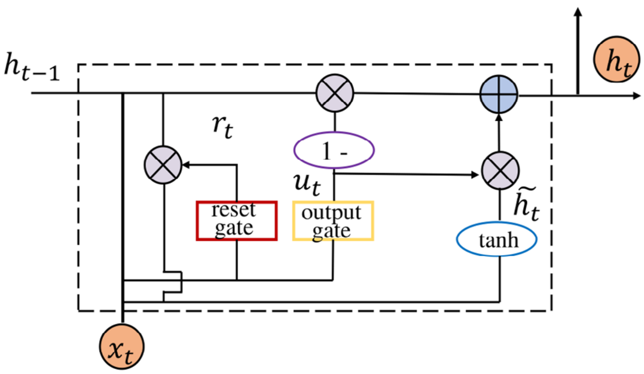


FIGURE 5 GRU structure.

degree of ignoring the historical state information, while the update gate changes its update speed according to the amount of historical state information brought in.

4.3 | Sequence-to-Sequence

The Seq2Seq model is a variant of the RNN. Seq2Seq has an encoder and a decoder. The encoder is generally composed of LSTM or other models for extracting features.The architecture is mainly divided into three parts: the encoder that is the input to the model, the decoder, which is the output of the model, and the encoder state vector. It can be seen that the encoder part is stacked with LSTM cells, and each of the cells allow a single element from the sequence as input [31]. The overall Seq2Seq-LSTM structure is shown in Figure 6.

5 | RESULTS AND DISCUSSION

Due to traffic self-similarity, the data segment can be arbitrarily selected from the trace. In our case, we chose the initial

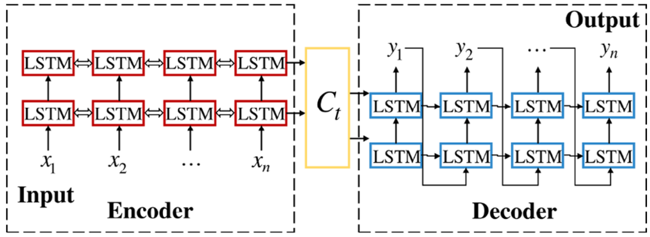


FIGURE 6 Seq2Seq-LSTM structure.

2880 data points from two datasets for experimentation. Within this subset, the first 1500 data points were allocated for training the traffic prediction model, the subsequent 690 data points served as the validation set, and the final 690 data points were designated as the test set.

Our experiments were conducted on the Ubuntu 22.0 operating system and TensorFlow 2.11.0 version.We initially employ Attention-LSTM, LSTM, GRU, and Seq2Seq models for satellite traffic prediction.The Attention-LSTM model consists of an LSTM layer with 64 hidden units, followed by a Lambda layer, a Dense layer, and a Fully Connected layer.Both the LSTM and GRU models comprise two layers, each with 64 and 32 hidden units, respectively. The Seq2Seq model includes two encoder and decoder layers, with 64 and 32 hidden units in each layer, respectively. To prevent overfitting of the model during training, an early stop mechanism has been added. The early stop mechanism is a regularisation tool to avoid overfitting on the training dataset. The early stop keeps track of the validation loss, and if the loss stops falling at successive epochs, the model training will be terminated. The model parameter settings are: batch_size: 32; earlyStopping: val_mse; epoch: 300; lr: 0.01; time step: 12.

The stability and results in the traffic forecasting is measured with a number of different metrics, we selected the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), as shown in Equations (8)–(10).

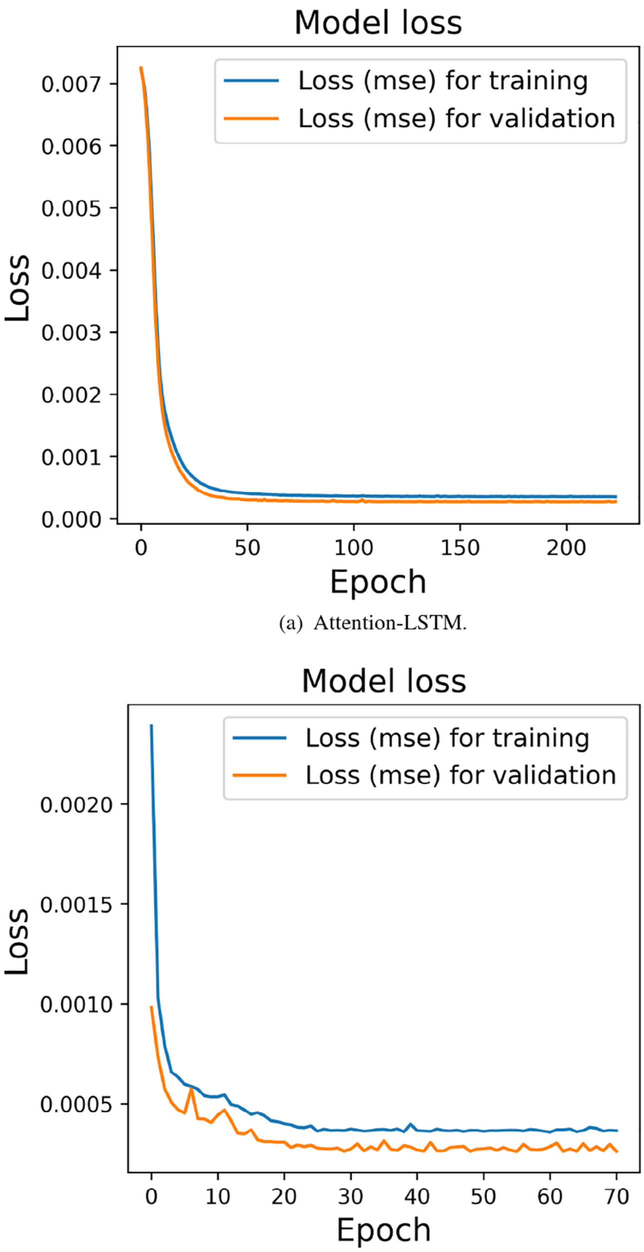
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}, \tag{8}$$

where $\hat{y}_i$ is the forecast value, y_i is the true value, n is the sample number. The greater the error, the greater the value.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|, \tag{9}$$

where $\hat{y}_i$ is the forecast value, y_i is the true value, n is the sample number. The greater the error, the greater the value.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|, \tag{10}$$



(a) Attention-LSTM.

(b) LSTM.

FIGURE 7 Model loss under ec_data.

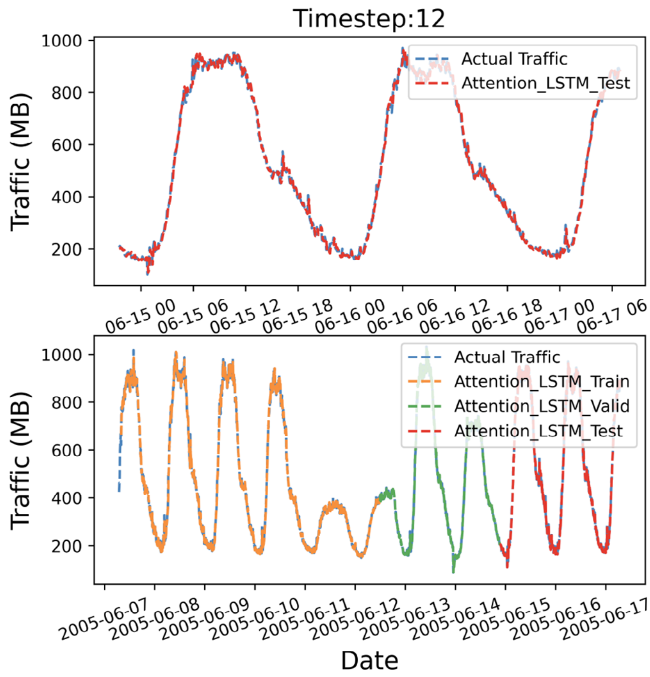


FIGURE 8 Attention-LSTM prediction result under ec_data.

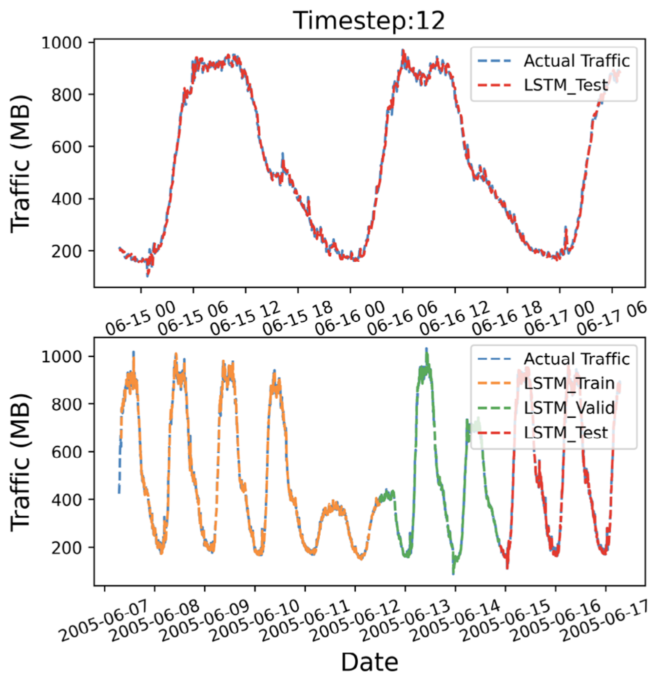
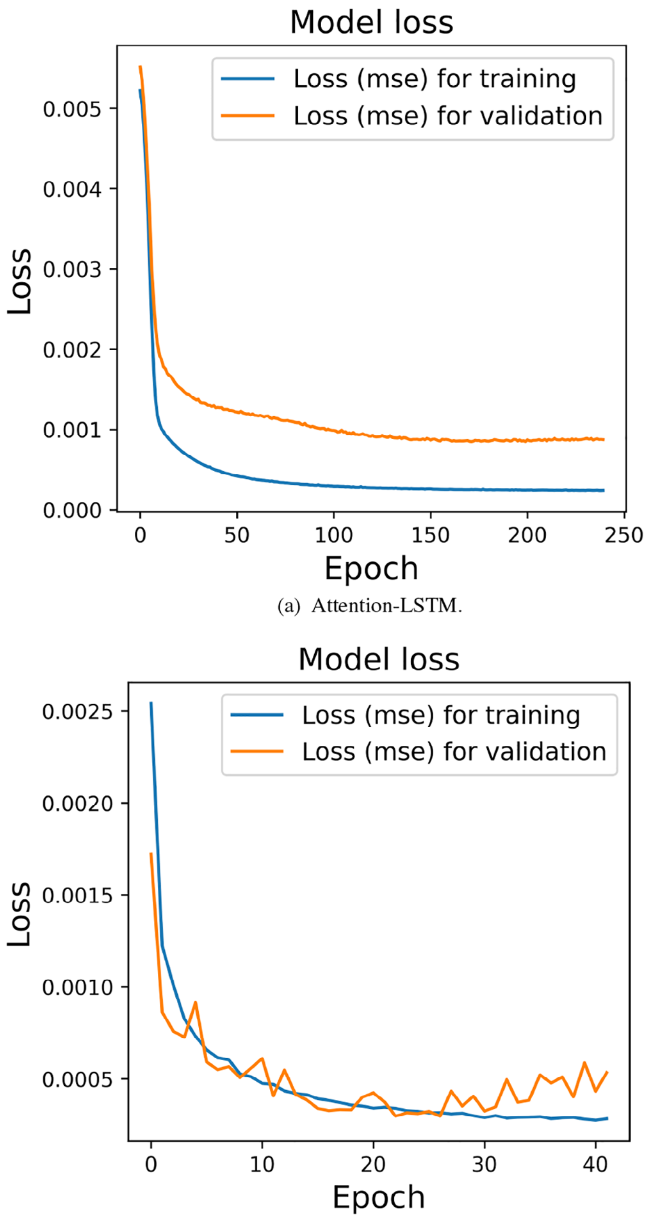


FIGURE 9 LSTM prediction result under ec_data.

where $\widehat{y}_i$ is the forecast value, y_i is the true value, n is the sample number. When the true value has data equal to 0, the equation is not available.

In Figure 7, we present the training and validation losses of the Attention-LSTM and LSTM models on ec_data. The Attention-LSTM exhibits a good overall fit and converges around round 30 with minimal divergence between them, whereas the LSTM model converges after roughly 20 rounds of slight oscillations, with minimal differentiation between training and validation losses. The upper portion displayed in Figure 8 represents a zoomed-in rendition of the model test set in the lower panel, offering a clearer visualization of how well the predicted results align with the actual traffic data (as depicted in the predicted results). Upon comparing Figures 8 and 9, we observe that the Attention-LSTM model excels in accurately predicting traffic extremes in contrast to the simple LSTM model, likely attributed to the incorporation of attention mechanisms, which facilitate the learning of pivotal information within the traffic flow.

In comparison to the ec_data, Figure 10 reveals that the validation loss of the Attention-LSTM model under uk_data



(a) Attention-LSTM.

(b) LSTM.

FIGURE 10 Model loss under uk_data.

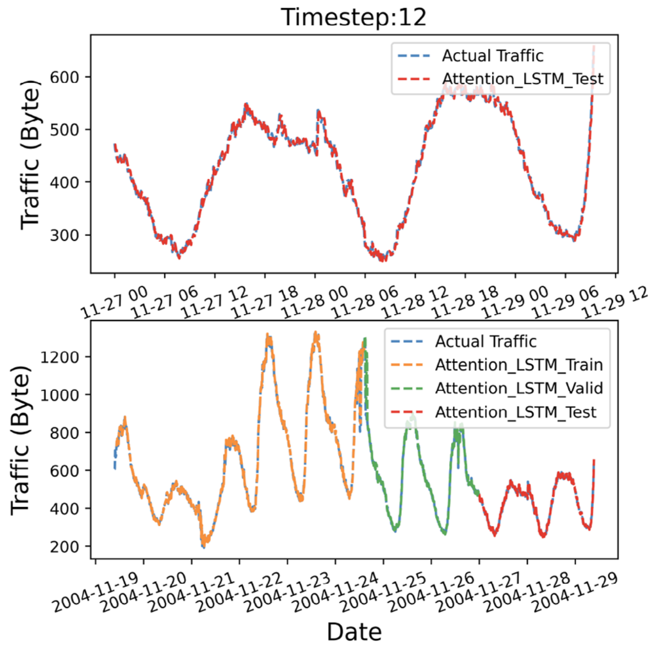


FIGURE 11 Attention-LSTM prediction result under uk_data.

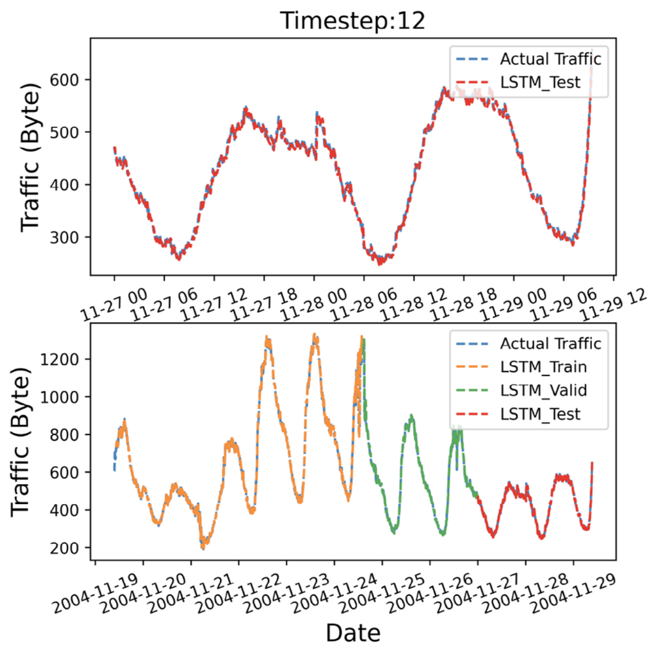


FIGURE 12 LSTM prediction result under uk_data.

experiences a rapid initial decrease, followed by convergence at a higher level than the training set, indicating overfitting. Conversely, the LSTM model exhibits wide-ranging fluctuations at convergence and gradually rises towards the end, suggesting biased predictions. In Figures 11 and 12, attention also contributes to obtaining crucial information about the traffic flow.

We have conducted ablation experiments on the LSTM variants of the GRU and Seq2Seq models, as well as comparative experiments using CNN models. The tables provided in Tables 3 and 4 present the RMSE, MAE, and MAPE of the five models for two datasets. Bold numbers highlight the best results of the five models at a time step of 12, effectively showcasing the outcomes of the five LEO satellite traffic prediction

TABLE 3 RMSE, MAE and MAPE for the five models under ec_data.

Model	RMSE	MAE	MAPE (%)
LSTM	18.419	13.478	3.33
GRU	18.579	13.493	3.35
Seq2Seq	18.967	13.817	3.48
CNN	24.445	18.12	4.52
Attention-LSTM	**18.351**	**13.371**	**3.31**

TABLE 4 RMSE, MAE and MAPE for the five models under uk_data.

Model	RMSE	MAE	MAPE (%)
LSTM	8.987	6.845	1.67
GRU	9.060	6.897	1.68
Seq2Seq	9.835	7.420	1.81
CNN	13.718	10.146	2.46
Attention-LSTM	**8.828**	**6.601**	**1.61**

TABLE 5 Diebold-Mariano test.

Ec_data	LSTM	GRU	CNN
Attention-LSTM	−0.5124	−0.9276	−9.6419
LSTM		−1.0984	−9.1671
GRU			−8.9997
Uk_data	**LSTM**	**GRU**	**CNN**
Attention-LSTM	−3.1519	−2.3294	−12.0753
LSTM		0.1628	−10.713
GRU			−11.0244

models. First, as mentioned earlier, both the LSTM and its variant model GRU consist of two layers, with 64 and 32 hidden units in each layer, respectively. Consequently, the variance in prediction results between the two traffic models is not extremely significant. However, the Seq2Seq model may encounter challenges during training due to the issue of gradient vanishing or explosion when handling time series data, leading to prediction results that are not as favorable as those produced by conventional LSTM and GRU models. Furthermore, the CNN prediction model performs the poorest, primarily because CNN fails to consider the sequential dependencies between output labels. On the other hand, the attention mechanism is capable of capturing dependencies and identifying crucial information among flows when confronted with lengthy sequence data, thereby yielding the most accurate prediction results.

The most straightforward approach to determining which model yields better flow prediction results is to compare the models based on the aforementioned prediction performance evaluation metrics and then select the prediction model with the lowest error. However, additional methods are necessary to assess the significance of differences between the prediction models. The Diebold-Mariano (DM) test is a statistical method used to evaluate the predictive ability of two or more forecasting models, based on their prediction error sequences. By employing a series of statistical tests, the DM test can determine whether there is a significant difference in predictive ability between the models under consideration. This makes it a useful tool for evaluating the predictive performance of different satellite traffic forecasting models, as defined below.

$$DM_{12} = \frac{\bar{d}_{12}}{\sigma_{d_{12}}}, \tag{11}$$

$$d_{12} = \frac{1}{N}\sum_{i=1}^{N}\left(\left(e_{it+1}^{(1)}\right)^2 - \left(e_{it+1}^{(2)}\right)^2\right), \tag{12}$$

where N represents the traffic data, d_{12} is the out-of-sample mean square error difference between the two prediction models, $\bar{d}_{12}$ and $\sigma_{d_{12}}$ denote the mean and standard deviation of d_{12}, respectively, and $((e_{it+1}^{(1)})^2$ and $(e_{it+1}^{(2)})^2$ are the prediction errors of the two prediction models for flow i, respectively.

If the value of DM is less than 0, it indicates that model 1 performs better than model 2. Table 5 provides the DM test results for each predictive model. Overall, all models demonstrate substantial significance when compared to CNN models. Among them, the Attention-LSTM model consistently exhibits superior predictive capability on both datasets. Its DM statistics are all below the critical value in comparison to the other three traffic forecasting models. This indicates that the Attention-LSTM model is well-suited for accurately predicting LSNs traffic.

6 | CONCLUSIONS

Satellite traffic serves as a critical indicator of the overall operational status of the satellite network. Accurate traffic prediction models are instrumental in resource allocation planning and network topology design. This paper introduces a self-similar traffic prediction approach for LEO satellite networks based on an improved LSTM model. Through an analysis of the characteristics and attributes of satellite network traffic, we conduct experiments utilizing Attention-LSTM and various LSTM-based traffic prediction models, applying them to two distinct traffic datasets. The experimental results demonstrate that the improved LSTM model exhibits superior performance in predicting self-similar traffic data. Moreover, the application of the Diebold-Mariano test underscores the significance of the model in comparison to other models. In conclusion, the proposed satellite traffic prediction method is well-suited for self-similar traffic characteristics and lays a solid foundation for future satellite network traffic planning.

AUTHOR CONTRIBUTIONS

Yan Zhang: Methodology; writing—original draft. **Yong Wang**: Methodology; resources; software; supervision. **Haotong Cao**: Methodology; resources; software; supervision. **Yihua Hu**: Funding acquisition; project administration; resources; software; supervision. **Zhi Lin**: Funding acquisition. **Kang An**: Writing—review and editing. **Dong Li**: Writing—review and editing.

ACKNOWLEDGEMENTS

This work was funded by the National Natural Science Foundation of China (62201592, 61901502 and 61671454), the Research Plan Project of NUDT (ZK21-33), the Young Elite Scientist Sponsorship Program of CAST (2021-JCJQ-QT-048), the Macau Young Scholars Program (AM2022011) and the

Postgraduate Scientific Research Innovation Project of Hunan Province (No. CX20220007).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data for this paper can be found in the references.

ORCID

Yan Zhang https://orcid.org/0000-0002-3483-1385
Yihua Hu https://orcid.org/0000-0002-1777-0242
Zhi Lin https://orcid.org/0000-0003-0011-7383
Kang An https://orcid.org/0000-0003-3635-3070

REFERENCES

1. Moscholios, I.D., Vassilakis, V.G., Sarigiannidis, P.G., Sagias, N.C., Logothetis, M.D.: An analytical framework in leo mobile satellite systems servicing batched poisson traffic. IET Commun. 12(1), 18–25 (2018)
2. Liu, J., Shi, Y., Fadlullah, Z.M., Kato, N.: space-air-ground integrated network: A survey. IEEE Commun. Surv. Tutorials 20(4), 2714–2741 (2018)
3. Del Portillo, I., Cameron, B.G., Crawley, E.F.: A technical comparison of three low earth orbit satellite constellation systems to provide global broadband. Acta Astronaut. 159, 123–135 (2019)
4. He, Y., Xiao, Y., Zhang, S., Jia, M., Li, Z.: Direct-to-Smartphone for 6G NTN: Technical Routes, Challenges, and Key Technologies. IEEE Network. 38(4), 128–135 (2024)
5. He, Y., Liu, Y., Jiang, C., Zhong, X.: Multiobjective anti-collision for massive access ranging in MF-TDMA satellite communication system. IEEE Internet Things J. 9(16), 14655–14666 (2022)
6. He, Y., Sheng, B., Yin, H., Yan, D., Zhang, Y.: Multi-objective deep reinforcement learning based time-frequency resource allocation for multi-beam satellite communications. China Commun. 19(1), 77–91 (2022)
7. Ma, R., Yang, W., Guan, X., Lu, X., Song, Y., Chen, D.: Covert mmWave communications with finite blocklength against spatially random wardens. IEEE Internet Things J. 11(2), 3402–3416 (2024)
8. Lin, Z., Niu, H., An, K., et al.: Refracting ris-aided hybrid satellite-terrestrial relay networks: Joint beamforming design and optimization. IEEE Trans. Aerosp. Electron. Syst. 58(4), 3717–3724 (2022)
9. Kozhaya, S.E., Haidar-Ahmad, J.A., Abdallah, A.A., Kassas, Z.M., Saab, S.S.: Comparison of neural network architectures for simultaneous tracking and navigation with leo satellites. In: Proceedings of the 34th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2021), pp. 2507–2520. The Institute of Navigation, Manassas, VA (2021)
10. Ma, R., Yang, W., Tao, L., Lu, X., Xiang, Z., Liu, J.: Covert communications with randomly distributed wardens in the finite blocklength regime. IEEE T. Veh. Technol. 71(1), 533–544 (2021)
11. Ma, R., Yang, W., Shi, H., Lu, X., Liu, J.: Covert communication with a spectrum sharing relay in the finite blocklength regime. China Commun. 20(4), 195–211 (2023)
12. Mokari, N., Hajipour, P., Mohammadi, L., Sardrood, P.S., Kashani, Z.G.: Resource allocation for non-delay-sensitive satellite services using adaptive coding and modulation–multiple-input and multiple-output–orthogonal frequency division multiplexing. IET Commun. 10(3), 309–315 (2016)
13. Katris, C., Daskalaki, S.: Comparing forecasting approaches for internet traffic. Expert Syst. Appl. 42(21), 8172–8183 (2015)
14. Li, J., Fu, D., Yuan, Q., et al.: A traffic prediction enabled double rewarded value iteration network for route planning. IEEE Trans. Veh. Technol. 68(5), 4170–4181 (2019)
15. Kim, H.W., Lee, J.H., Choi, Y.H., Chung, Y.U., Lee, H.: Dynamic bandwidth provisioning using arima-based traffic forecasting for mobile wimax. Comput. Commun. 34(1), 99–106 (2011)
16. Liu, P., Chen, H., Wei, S., Li, L., Zhu, Z.: Hybrid-traffic-detour based load balancing for onboard routing in leo satellite networks. China Commun. 15(6), 28–41 (2018)
17. Song, L., Bondon, P., Cao, Y., Cheng, Q.: A time-varying farima model for internet traffic. In: 2008 Congress on Image and Signal Processing, vol. 5, pp. 83–87. IEEE, Piscataway (2008)
18. Gao, B., Zhang, Q.Y., Liang, Y.S., Liu, N.N., Huang, C.B., Zhang, N.T.: Predicting self-similar networking traffic based on emd and arma. J. China Inst. Commun. 32(4), 47–56 (2011)
19. Bie, Y., Wang, L., Tian, Y., Hu, Z.: A combined forecasting model for satellite network self-similar traffic. IEEE Access 7, 152004–152013 (2019)
20. Tian, Z.: Network traffic prediction method based on wavelet transform and multiple models fusion. Int. J. Commun. Syst. 33(11), e4415 (2020)
21. Liang, Y., Qiu, L.: Network traffic prediction based on svr improved by chaos theory and ant colony optimization. Int. J. Futur. Gener. Co. 8(1), 69–78 (2015)
22. Na, Z., Pan, Z., Liu, X., Deng, Z., Gao, Z., Guo, Q.: Distributed routing strategy based on machine learning for leo satellite network. Wirel. Commun. Mob. Com. 2018 (2018)
23. Li, N., Hu, L., Deng, Z.L., Su, T., Liu, J.W.: Research on gru neural network satellite traffic prediction based on transfer learning. Wirel. Pers. Commun. 118, 815–827 (2021)
24. Na, Z., Liu, Y., Cui, Y., Guo, Q.: Research on aggregation and propagation of self-similar traffic in satellite network. Int. J. Hybrid Inf. Technol. 8(1), 325–338 (2015)
25. Cortez, P., Rio, M., Rocha, M., Sousa, P.: Internet traffic forecasting using neural networks. In: The 2006 IEEE International Joint Conference on Neural Network Proceedings, pp. 2635–2642. IEEE, Piscataway (2006)
26. Cortez, P., Rio, M., Rocha, M., Sousa, P.: Multi-scale internet traffic forecasting using neural networks and time series methods. Expert Syst. 29(2), 143–155 (2012)
27. Wang, Y., Vijayakumar, P., Gupta, B.B., Alhalabi, W., Sivaraman, A.: An improved entity recognition approach to cyber-social knowledge provision of intellectual property using a crf-lstm model. Patt. Recogn. Lett. 163, 145–151 (2022)
28. Zhao, L., Hu, Y., Wang, B., Jiang, X., Liu, C., Zheng, C.: A surrogate-assisted evolutionary algorithm based on multi-population clustering and prediction for solving computationally expensive dynamic optimization problems. Expert Syst. Appl. 223, 119815 (2023)
29. Wang, D., Mao, K.: Multimodal object classification using bidirectional gated recurrent unit networks. In: 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), pp. 685–690. IEEE, Piscataway (2018)
30. Zhao, L., Hu, Y., Yang, X., Dou, Z., Kang, L.: Robust multi-task learning network for complex lidar point cloud data preprocessing. Expert Syst. Appl. 237, 121552 (2024)
31. Masood, Z., Gantassi, R., Choi, Y.: A multi-step time-series clustering-based seq2seq lstm learning for a single household electricity load forecasting. Energies 15(7), 2623 (2022)

How to cite this article: Zhang, Y., Wang, Y., Cao, H., Hu, Y., Lin, Z., An, K., Li, D.: Self-similar traffic prediction for LEO satellite networks based on LSTM. IET Commun. 19, e12863 (2025). https://doi.org/10.1049/cmu2.12863