Dean&Francis
ISSN 2959-6157

# Research and Implementation of Network Traffic Prediction Algorithm Based on Informer

**Feiyang Xiao**

Department of Communication Engineering, Wuhan University of Technology, Wuhan, 430070, China
Corresponding author: feiyang1919@outlook.com

**Abstract:**

As next-generation mobile networks like 5G and 6G proliferate and internet connectivity becomes more accessible, network traffic prediction plays a crucial role in network management. Excellent traffic prediction results can enhance network management efficiency and improve network bandwidth utilization, making efficient traffic prediction essential. Traditional models primarily focus on the temporal characteristics of network traffic while neglecting its spatial characteristics, resulting in poor accuracy and stability in traffic prediction. However, the development of deep learning has provided new solutions for predicting network traffic. This paper conducts an in-depth study of the Transformer time series algorithm, identifies its shortcomings in network traffic data prediction, and introduces the ProbSparse self-attention mechanism into the model. Based on this, network traffic prediction algorithms can be implemented using Informer. This study utilizes network traffic usage data from the University of Massachusetts Amherst and its Amherst campus. Four sets of experiments were conducted under different prediction step conditions to compare the prediction results with real traffic data across four metrics: MAP, RMSE, MAPE, and $R^2$, thereby evaluating whether the Informer model can maintain prediction accuracy over extended periods.

**Keywords:** Network Traffic Prediction; Informer; Evaluation Indicators; Feature Extraction.

## 1. Introduction

The onset of the information age has accelerated computer development, network, and communication technologies [1]. Among these technologies, one of the 21st century's fastest-growing sectors is the Internet [2], and global networks have become an indispensable part of human society. With the development of smart devices, everyone can now easily communicate globally via the Internet. According to

statistics, as of now, the number of Internet users worldwide has exceeded billions, demonstrating that the Internet is developing at an increasingly rapid pace, with an ever-growing number of people joining its user base [3].
In order to overcome Transformers' shortcomings in long-sequence time series prediction tasks, Zhou and colleagues introduced the Informer neural network model [4]. In addition to offering a novel embedding technique that combines time, position, and data information, Long-range dependencies in long-sequence time series are effectively captured by the model through the use of the ProbSparse self-attention mechanism. With the goal to enhance the ability to record long-range dependency data, Jacek et al. [5] created an artificial neural network model using a sparse self-attention mechanism that was based on Informer. This paper aims to study and implement a network traffic prediction algorithm based on the Informer model to achieve accurate predictions of network traffic using this model.

## 2. Related work

### 2.1 Informer Model

The essence of the Informer model is the attention mechanism combined with the Transformer model [6]. The Informer model's main concept is to use the self-attention mechanism to examine the input sequence in order to identify long-term dependencies in it, and then utilize the Transformer's encoder-decoder structure for prediction.
The three most important components of the Informer model are: the Encoder module, the Decoder module, and the Output module. The Encoder module has a relatively simple structure, with each layer in the module being identical. Each layer includes an attention mechanism and a convolutional layer [7]. As previously mentioned, Long-range dependencies in the data are captured by the attention mechanism. Global features are successfully extracted from the data by this method. Local features are captured by the convolutional layer, and when the two are combined, the model can simultaneously collect global and local characteristics, greatly enhancing its prediction ability.
Additionally, the decoder is made up of several identical layers, each of which has a convolutional layer and a self-attention mechanism [8]. In order to produce an output sequence based on the encoder's hidden representation, the decoder must efficiently use the encoder's data in order to produce precise predictions. To this end, the decoder introduces an attention mechanism called "Encoder-DecoderAttention" at each position to perform a weighted combination of the encoder's hidden representation.

## 3. Methodology

As shown in Fig. 1, the output layer is used to convert the decoder's output into the final prediction result. Typically, the output layer includes a fully connected layer to map the decoder's hidden representation to the prediction target space, followed by an activation function (such as a linear function or softmax function) to obtain the final prediction result [9].

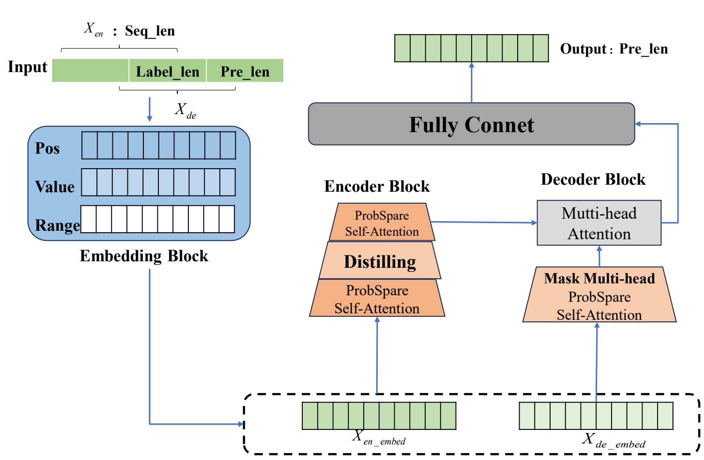


**Fig. 1 Informer overall architecture diagram**

### 3.1 Data and Evaluation Indicators

The dataset used in this experiment was collected between 07:00 on July 6, 2015, and 13:55 on July 28, 2015, from the University of Massachusetts in Massachusetts, USA (denoted as mes).
Mean Absolute Error (MAE), Root Mean Square Error

ISSN 2959-6157

(RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R2) are the four metrics most frequently used to assess model performance in experiments.

MAE: The average of the absolute values of the deviations between the arithmetic mean and each individual observation is known as the mean absolute error.

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

In this formula $y_i$ represents the actual network traffic value; $\hat{y}_i$ represents the predicted network traffic value; and n represents the number of prediction data points.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100\% \tag{2}$$

In this formula, projected value is denoted by $\hat{y}_t$, the actual value by $y_t$, and the quantity of data utilized for evaluation by $n$.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (f_i - \hat{f}_i)^2}{\sum_{i=1}^{n} (f_i - \bar{F})^2} \tag{3}$$

The implementation process of the algorithm consists of data loading and time processing. In this procedure, the number of model training iterations and the length of the input and output data are predetermined. Before being separated into training, validation, and test sets, the data is first read and standardized. Then, the model parameters are adjusted, followed by model training, result saving, and output [10].

# 4. Results

In this experiment, mes network traffic data was selected as experimental data. Under this traffic flow variation scenario, the information model's prediction performance is examined for various prediction steps.

At various forecast lead periods, the informer model's predictive ability was examined for this traffic data. The input data lead time was set to 48, traffic data from 48 historical time points, while the output lead times were divided into 1, 12, 24, and 48, predicting traffic results for the next 1, 12, 24, and 48 time points. The prediction results are shown in the Fig. 2.

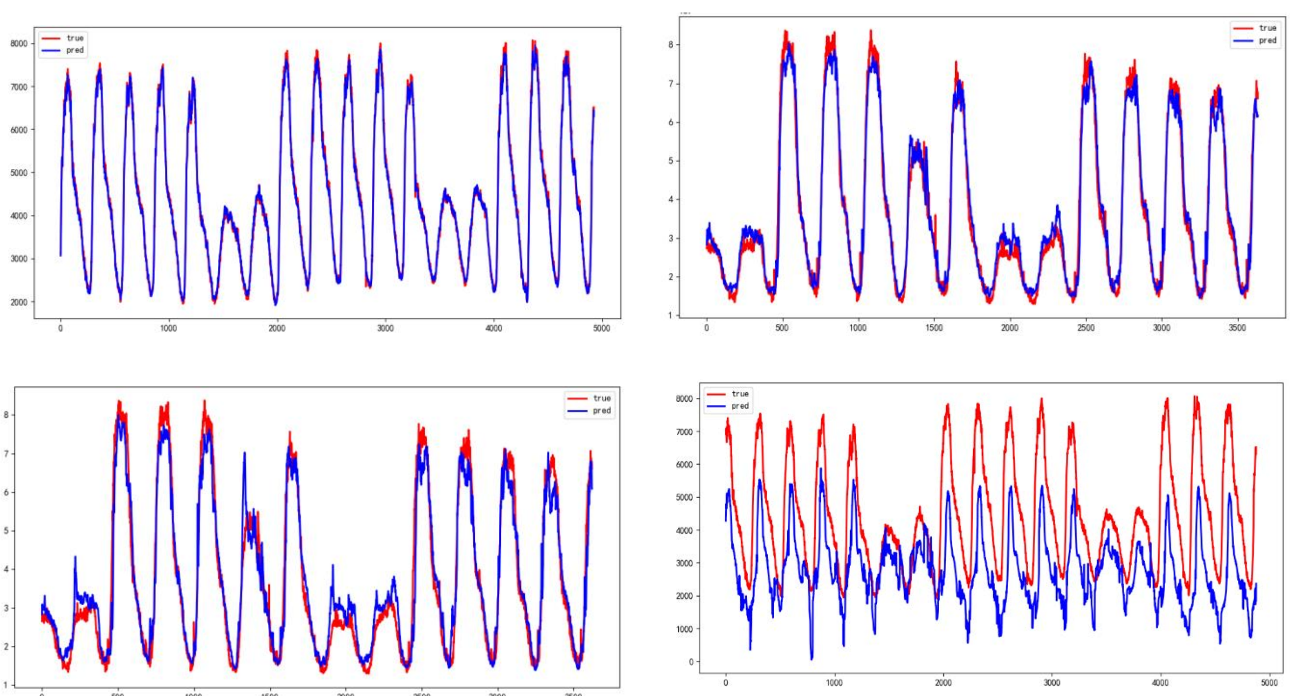


**Fig. 2 Predicted results for step lengths of 1, 12, 24, and 48**

The aforementioned graphic illustrates how the red and blue curves significantly overlap when single-step forecasting is used with a step size of 1. When predicting with a step size of 12, the red and blue curves also overlap to a quite high extent. But the forecasting error is larger compared to the case with a step size of 1. In the case of a step size of 24, the blue curve and the red curve show significant discrepancies in many areas, resulting in a further decrease in prediction accuracy. In the case of a step size of 48, the blue curve and the red curve show numerous discrepancies in many areas. Based on the above four predictions, it can be inferred that as the prediction step size increases, the error becomes larger, and the corresponding accuracy decreases.

**Table 1. Comparison of error rates of the Informer model on the mes dataset**

| Dataset | Inputlength | Prediction stride | MAE | RMSE | MAPE | $R^2$ |
|---|---|---|---|---|---|---|
| mes | 48 | 1 | 2.41E+08 | 3.51E+08 | 7.545542 | 0.970786 |
| | 48 | 12 | 2.75E+08 | 3.93E+08 | 8.412892 | 0.963464 |
| | 48 | 24 | 3.93E+08 | 5.11E+09 | 11.87896 | 0.938362 |
| | 48 | 48 | 5.18E+08 | 7.04E+08 | 18.54684 | 0.883555 |

As shown in the Table 1, the longer the predicted step length, the corresponding MAE and RMSE average absolute percentage error all increase accordingly, so the coefficient of determination $R^2$ (accuracy) gradually decreases. It can be concluded that, under the condition of unchanged input length, when the prediction step size is 1, the accuracy is very high and is basically consistent with the actual results. Correspondingly, the longer the input step size, the lower the accuracy [11].

## 5. Conclusion

In this paper, a network traffic prediction algorithm was implemented and investigated using the Informer model. By analyzing and modeling network traffic data, we can gain a deeper understanding of network traffic behavior and use deep learning technology to achieve accurate traffic predictions. Throughout the research process, we went through many steps, including data preparation, model construction, training and validation, and result evaluation, and achieved significant progress and results.

First, we conducted a detailed analysis and preprocessing of the network traffic data. We extracted key features from the data and performed preprocessing operations such as data normalization and sequence length adjustment to ensure data quality and applicability. This laid a solid foundation for subsequent model training and prediction.

Second, we selected the Informer model as the base model for network traffic prediction. The Informer model possesses strong time series modeling capabilities, effectively capturing spatio-temporal relationships and dynamic changes in the data. We adjusted the model parameters and hyperparameters based on actual conditions to maximize the model's performance and generalization capabilities.

During the model training and validation phase, we conducted rigorous experiments to ensure the model's robustness and reliability. Through multiple iterations of training on the training and validation sets and fine-tuning based on the validation set's performance metrics, we gradually improved the model's predictive performance and achieved satisfactory results.

Finally, after model training was completed, we conducted a comprehensive evaluation and testing of the model. The model's predictive performance was thoroughly analyzed and compared using a number of widely used assessment criteria, including RMSE and MAE. The experimental results denote that the network traffic prediction algorithm based on the Informer model achieved satisfactory predictive performance to a certain extent, demonstrating significant advantages and improvements compared to traditional methods.

In summary, this study has achieved certain results in the implementation of a network traffic prediction algorithm based on the Informer model, providing useful references and insights for future related research and practice. However, we are also aware that there are still some challenges and room for improvement in practical applications, such as further optimizing the model structure, improving data preprocessing methods, and enhancing the model's generalization capabilities.

## References

[1] U. Cisco, “Cisco annual internet report (2018–2023) white paper,” Cisco, San Jose, CA, USA, White Paper C11, 2020, vol. 10, no. 1, pp. 1–35.

[2] Wang, C.; Chen, Y.; Zhang, S.; Zhang, Q. Stock market index prediction using deep Transformer model. Expert Syst. Appl. 2022, 208, 118128.

[3] Aseeri AO (2023) Effective short-term forecasts of saudi stock price trends using technical indicators and large-scale multivariate time series. Peerj Comput Sci 9:e1205.

[4] Vollebregt, E.; Six, K.; Polach, O. Challenges and progress in the understanding and modelling of the wheel–rail creep forces. Veh. Syst. Dyn. 2021, 59, 1026–1068.

[5] Limperis J, Tong W, Hamza-Lup F, Li L (2023) Pm2.5 forecasting based on transformer neural network and data

ISSN 2959-6157

embedding. Earth Sci Inform 16(3):2111–2124.
[6] Gao BY, Wang RR, Lin CJ, et al., 2021. TBM penetration rate prediction based on the long short-term memory neural network. Underground Space, 6(6):718–731.
[7] Li, Z.; Rao, Z.; Pan, L.; Xu, Z. MTS-Mixers: Multivariate Time Series Forecasting via Factorized Temporal and Channel Mixing. arXiv 2023, arXiv:2302.04501.
[8] Alarrouqi RA, Bayhan S, Al-Fagih L. An assessment of the energy performance of battery-electric buses in hot environments. Sustain Energy Grids Networks.2024;38:101352.
[9] Zang ZS, Li ZH, Zhan X, Kong XG, Gu ZJ, Yin S ,Niu Y (2024) The multi-dimensional joint response of the electric potential time–frequency-field in the fracture process of flawed coal under dynamic load. Theor Appl Fract Mech 131:104415.
[10] Wu, J.M.T.; Li, Z.; Herencsar, N.; Vo, B.; Lin, C. A graph-based CNN-LSTM stock price prediction algorithm with leading indicators. Multimed. Syst. 2023, 29, 1751–1770.
[11] Jin, J.; Jin, Q.; Chen, J.; Wang, C.; Li, M.; Yu, L. Prediction of the Tunnelling Advance Speed of a Super-Large-Diameter Shield Machine Based on a KF-CNN-BiGRU Hybrid Neural Network. Measurement 2024, 230, 114517.