

Article

DBSTGNN-Att: Dual Branch Spatio-Temporal Graph Neural Network with an Attention Mechanism for Cellular Network Traffic Prediction

Zengyu Cai ¹, Chunchen Tan ¹, Jianwei Zhang ^{2,3,*}, Liang Zhu ¹  and Yuan Feng ¹

¹ School of Computer Science and Technology, Zhengzhou University of Light Industry, Zhengzhou 450000, China; czy@zzuli.edu.cn (Z.C.); 332107040615@email.zzuli.edu.cn (C.T.); lzhu@zzuli.edu.cn (L.Z.); mailfengy@163.com (Y.F.)

² School of Software Engineering, Zhengzhou University of Light Industry, Zhengzhou 450000, China

³ ZZULI Research Institute of Industrial Technology, Zhengzhou 450003, China

* Correspondence: mailzjw@163.com

Abstract: As network technology continues to develop, the popularity of various intelligent terminals has accelerated, leading to a rapid growth in the scale of wireless network traffic. This growth has resulted in significant pressure on resource consumption and network security maintenance. The objective of this paper is to enhance the prediction accuracy of cellular network traffic in order to provide reliable support for the subsequent base station sleep control or the identification of malicious traffic. To achieve this target, a cellular network traffic prediction method based on multi-modal data feature fusion is proposed. Firstly, an attributed K-nearest node (KNN) graph is constructed based on the similarity of data features, and the fused high-dimensional features are incorporated into the graph to provide more information for the model. Subsequently, a dual branch spatio-temporal graph neural network with an attention mechanism (DBSTGNN-Att) is designed for cellular network traffic prediction. Extensive experiments conducted on real-world datasets demonstrate that the proposed method outperforms baseline models, such as temporal graph convolutional networks (T-GCNs) and spatial-temporal self-attention graph convolutional networks (STA-GCNs) with lower mean absolute error (MAE) values of 6.94% and 2.11%, respectively. Additionally, the ablation experimental results show that the MAE of multi-modal feature fusion using the attributed KNN graph is 8.54% lower compared to that of the traditional undirected graphs.

Keywords: cellular network traffic prediction; deep learning; graph neural network; multi-modal feature fusion; attention mechanism



Citation: Cai, Z.; Tan, C.; Zhang, J.; Zhu, L.; Feng, Y. DBSTGNN-Att: Dual Branch Spatio-Temporal Graph Neural Network with an Attention Mechanism for Cellular Network Traffic Prediction. *Appl. Sci.* **2024**, *14*, 2173. <https://doi.org/10.3390/app14052173>

Academic Editor: Christos Bouras

Received: 15 January 2024

Revised: 24 February 2024

Accepted: 1 March 2024

Published: 5 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the advent of the information age and the rapid advancements in 4th generation mobile communication technology (4G) and 5th generation mobile communication technology (5G), the network has become an indispensable tool and medium for accessing information in today's society [1]. At the same time, the constant updates and improvements in various intelligent devices and the significant increase in the number of mobile network users have posed significant challenges in terms of configuring network resources and constructing network infrastructure. With the ongoing development of 5G and the accelerated progress of 6th generation mobile communication technology (6G), this remarkable growth trajectory will persist [2]. In recent years, the domain of network traffic has witnessed extensive research in areas such as network traffic classification, network traffic anomaly detection, and network traffic prediction [3–5]. These studies have provided invaluable insights for the advancement of the network field.

However, in comparison to image recognition and natural language processing, the datasets published in the field of cellular network traffic prediction are relatively limited

in diversity [6]. The quality of the datasets is low and they contain a significant amount of noise. To avoid compromising the experimental results with low-quality data, certain researchers have resorted to using private domain data. While this approach can enhance the model's efficiency, it hinders the ability to compare it to other models and advance the entire field. Furthermore, integrating multiple types of data through feature fusion can enhance the accuracy of predicting cellular network traffic [7]. Previous studies have successfully utilized considerable information, such as the number of base stations, weather data, and points of interest (POIs), to achieve remarkable results [8]. However, the traditional undirected graph overlooks the local similarities between samples and is more suitable for depicting global connection relationships. As the data similarity between adjacent regions in cellular network traffic is higher, the undirected graph fails to meet the demands of predicting such traffic. In terms of prediction methods, deep learning models, particularly convolutional neural networks (CNNs) and long short-term memory networks (LSTMs), have been extensively utilized in cellular network traffic prediction. Variants based on CNNs and LSTMs have also presented excellent effects. However, with the proliferation of mobile devices, the scale of network traffic is expanding. Some existing models fail to consider the spatial and geographical characteristics of traffic, resulting in issues of low accuracy and a lack of interpretability when handling long-term predictions. Consequently, there is an urgent need to design a long-term prediction model for cellular network traffic that incorporates multi-modal features.

This paper addresses three predicaments: (1) The effective fusion of multi-modal features; (2) enhancing the accuracy of predictions; and (3) verifies the effectiveness and feasibility of the proposed method in the real world. With the intention of resolving these issues, this paper makes the following contributions:

- An attributed KNN graph is constructed. According to the similarity of nodes, the most similar K nodes are extracted and input into the model, which avoids the low prediction accuracy caused by the uneven spatial and temporal distribution of cellular traffic. In addition, the attributed KNN graph also combines multi-modal features, such as POIs, to provide more information for the model.
- A prediction model called DBGTNN-Att is developed. The model utilizes spatial graph convolution (SGC) to process the multi-modal features present in the KNN graph and subsequently feeds the processed features into dual branch-gated convolution with an attention mechanism (DBGCA) for time-series prediction to capture the complex relationships in the data.
- We evaluate our proposed DBSTGNN-Att by comparing it to other baseline prediction models based on T-GCN or STA-GCN using a real-world cellular network dataset in short-term, mid-term, and long-term prediction scenarios. We also discuss several exciting research opportunities for cellular traffic prediction in the future.

The rest of the article is organized as follows. Section 2 reviews the related work of cellular traffic prediction. Section 3 shows the structure of DBSTGNN-Att, including the attributed KNN graph of multi-modal features and the functions of each component within the model. Section 4 is the experimental part, by comparing the various baseline models to prove the effectiveness of the proposed method. Finally, the conclusion is drawn in Section 5.

2. Related Work

The prediction of cellular network traffic involves forecasting future cellular network traffic data through the analysis of the spatial-temporal distribution of known cellular traffic data. Over the past decade, deep learning techniques have gained widespread application in time-series prediction, including the prediction of vehicle flow and subway passenger flow [9–14]. Incorporating deep learning into time-series prediction has significantly contributed to the advancement of cellular network traffic prediction. This approach eliminates the need for the manual design of feature standards, reducing the influence of human factors on the experimental results and enhancing both operational efficiency and

prediction accuracy. As a result, we decided to utilize deep learning in cellular network traffic prediction to minimize resource consumption and enhance the prediction accuracy.

During the initial stages, a considerable number of studies utilized probability statistical approaches or probability distribution methods to predict cellular traffic, including models like auto-regression moving average (ARMA) and auto-regressive integrated moving average (ARIMA) [15–17]. For instance, Yang, H., et al. [18] proposed a cellular network traffic prediction model that integrates simulated annealing (SA), ARIMA, and a back-propagation neural network (BPNN). This model effectively exploited the potential of extracting both linear and nonlinear patterns from historical cellular traffic data, leading to an enhanced accuracy in predictions. However, the traditional regression models rely heavily on extensive pre-existing knowledge based on queuing theory and network traffic flow theory and exhibit a limited efficacy in handling intricate uncontrolled variables. With the increasing complexity of networks, these models often yield a subpar performance in actual cellular network traffic prediction tasks, resulting in a reduced accuracy.

With the ongoing progress of technology, the utilization of support vector machines (SVMs) and K-nearest neighbors (KNNs) has become widespread in the domain of cellular network traffic prediction [19,20]. These models convert the prediction problem into a linear partition problem, reducing the artificial impact on prediction results and enhancing the prediction accuracy to a certain degree. However, the drawbacks of these models are evident due to the nonlinear and non-stationary nature of network traffic itself. In scenarios with numerous data samples or high dimensions, a substantial amount of computing resources is required, which fails to meet the demands of long-term network traffic prediction.

With the progressive maturity of deep learning models, such as deep neural network (DNN), CNN, and LSTM [21,22], they have commenced to yield favorable outcomes in cellular network traffic prediction. Zhang, S. [23] employed particle swarm optimization (PSO) for optimizing the parameters of variable mode decomposition (VMD) and then reconstructing the cellular network traffic data. Subsequently, bidirectional long short-term memory network (BiLSTM) was utilized to carry out cellular traffic prediction. This approach circumvents the influence of manually configured parameters on the experimental outcomes and effectively mitigates the issue of inadequate feature extraction resulting from unidirectional learning. Nevertheless, one of the main drawbacks of these models is that they do not consider the spatial distribution characteristics of network traffic, but only focus on its time distribution in time series. Consequently, this often leads to a diminished prediction accuracy and a significant utilization of artificial resources.

In recent years, there has been a growing trend in utilizing models like graph neural network (GNN) and CNN that demonstrate proficiency in extracting spatial features for the prediction of cellular network traffic [24,25]. These models have proven to be effective in this domain. For instance, Shen, W., et al. [26] propose a time-wise attention-aided convolutional neural network (TWACNet) structure for cellular traffic prediction. In the proposed TWACNet, the time-wise attention mechanism is adopted to capture the long-range temporal dependencies of the cellular traffic data, and the convolutional neural network (CNN) is adopted to capture the spatial correlation. To effectively capture the correlation between time and space within network traffic, Wang, Z., et al. [27] introduced GNN as a tool for cellular network traffic prediction. This particular model is capable of extracting the spatio-temporal correlation and characteristics of intercellular traffic, leading to highly accurate predictions regarding network traffic patterns. Building upon the foundation of GNN, Zhou, X., et al. [28] proposed a transfer learning strategy. The incorporation of this strategy yields substantial benefits, such as preserved computing resources and enhanced model fitting speed, ultimately resulting in the reduced operational time of the model.

Therefore, this research introduces the concept of constructing a dual branch spatio-temporal graph neural network. This model incorporates the fusion of multi-modal features while simultaneously combining a self-attention mechanism to effectively combine and

process the multi-modal information obtained from cellular network traffic. Additionally, the proposed model exhibits improved interpretability, resulting in a heightened accuracy in prediction.

3. Methodology

In this section, the exposition primarily focuses on delineating the construction of the KNN graph, as well as elucidating the structure and function of each component within the DBSTGNN-Att model.

3.1. Attributed KNN Graph

In the existing research in the field of cellular traffic prediction, the fusion of multi-modal features often uses multiple undirected graphs to input into the neural network. Although this method can provide more abundant information for the model, it also improves the complexity and running time of the model. In addition, the method of expressing the correlation of each region by a 0.1 matrix cannot accurately extract the complex spatial and temporal distribution of cellular traffic, which is due to the large difference in the traffic distribution pattern of adjacent regions. Therefore, this paper selected the most similar K nodes for prediction by constructing the attributed KNN graph, and the multi-modal features were directly spliced after the historical time series. This approach not only reduced the complexity of the model, but also improved the prediction accuracy of the model.

The multi-modal features encompass comprehensive information pertaining to cellular traffic, encompassing the time of traffic occurrence, the volume of traffic at that specific time, the quantity of base stations (BSs), the quantity of points of interest (POIs), and social activities (SAs). The number of base stations reflects the peak value of cellular traffic in this region. No matter how large the traffic load demand in this region is, the peak value of traffic will not exceed the maximum bearing range of the base station. Points of interest (POI) within the region offer insights into the extent of population concentration, with areas containing a larger number of POIs indicating a higher degree of aggregation, as well as SAs. The correlation between multi-modal features and the distribution of cellular traffic, as well as the spatial distribution characteristics of cellular traffic, are depicted in Figure 1.

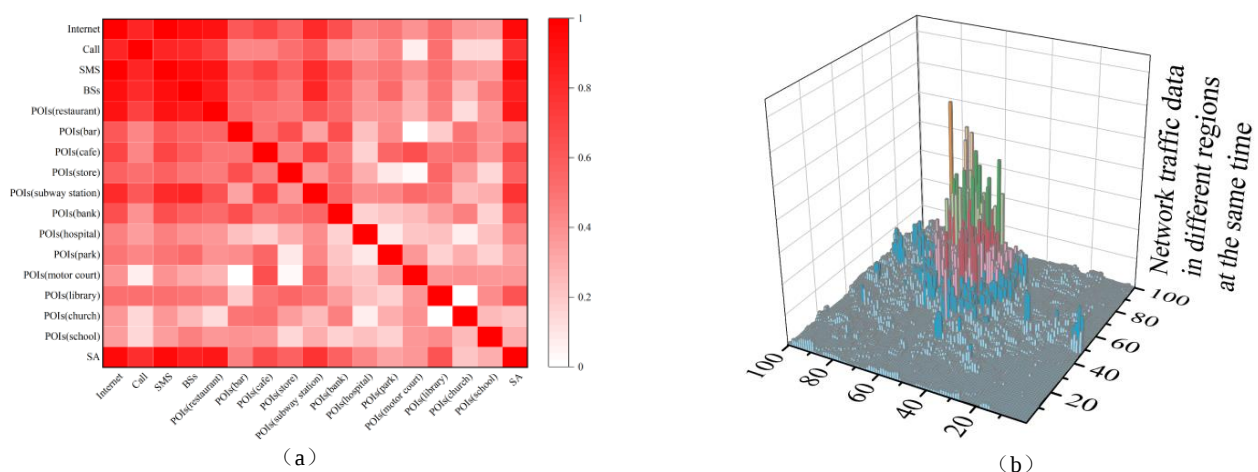


Figure 1. (a) The correlation between multi-modal features and the distribution of cellular traffic. (b) Spatial distribution of cellular network traffic (the data in (b) is from the dataset [29] used in this paper, which divides the Italian city of Milan into 10,000 regions to form a 100×100 square matrix).

Consequently, we established an attributed KNN graph by evaluating the similarity between cellular traffic in different regions. The specific procedure for transforming cellular traffic into a KNN graph with node attributes can be observed in Figure 2.

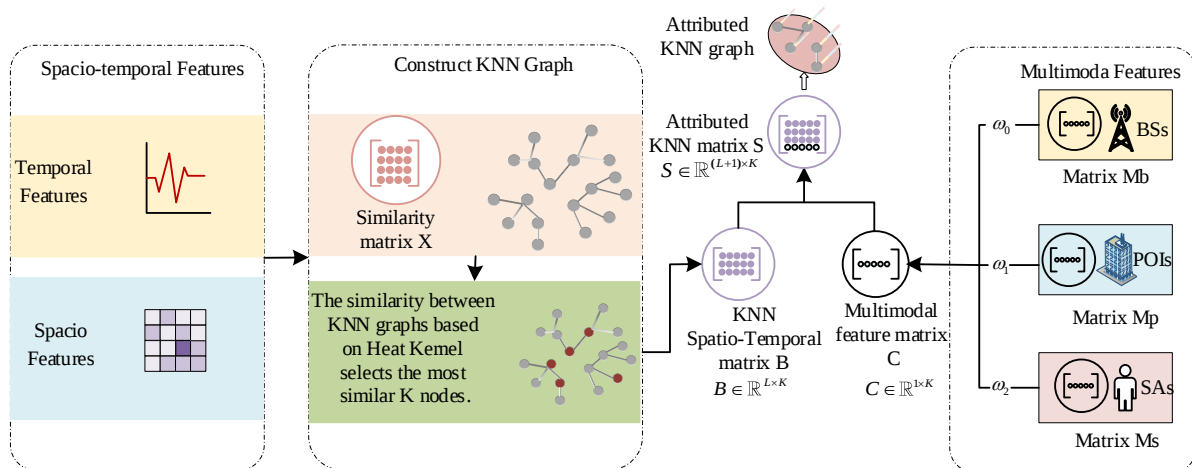


Figure 2. Attributed KNN graph construction.

The attributed KNN graph can be comprehended as a type of similarity matrix. The measurement of the distance between regions is based on the spatial attributes of the cellular traffic. Regions that are closer exhibit a higher similarity, whereas regions that are farther apart display lower levels of similarity, and in some cases, may even be considered insignificant. However, not all spatial distributions are related to distance. Some regions are very close to each other, but the traffic distribution patterns are completely different; some regions are far apart, but the traffic distribution patterns are very similar.

Therefore, this paper first selected a target node and obtained the similarity matrix X between the node and the other nodes by heat kernel [30] calculations. According to the similarity, K nodes with the closest attributes were selected from high to low. Each node represents the cellular traffic data in a region (for example, the Milan dataset used in this paper divided Milan into a 100×100 matrix; so, the number of nodes was 10,000). The cellular traffic data of these K nodes was extracted and rearranged into a $L \times K$ matrix, namely KNN spatio-temporal matrix B , where K is the number of the most similar K nodes and L is the length of the time series. Then, the number of BS, POIs, and SAs in the region corresponding to these K nodes was extracted to form three $1 \times K$ matrices, namely matrix M_b , matrix M_p , and matrix M_s . According to the correlation of the multi-modal features, as shown in Figure 1a, the matrix of each multi-modal feature was assigned the corresponding weights ω_0 , ω_1 , and ω_2 . The matrices M_b , M_p , and M_s were multiplied by the corresponding weights, and then the multi-modal feature matrix C was obtained by adding them, and it was spliced with the KNN spatio-temporal matrix B to obtain a $(L + 1) \times K$ matrix, that is, the matrix S .

The similarity between node i and node j calculated by heat kernel is shown in Equation (1).

$$X_{i,j} = \frac{1}{L} \sum_{t=1}^L e^{-\frac{|x_i - x_j|^2}{t}} \quad (1)$$

where t is the time parameter in the heat conduction equation.

According to Equation (1), if the target node is i , the closer the similarity $X_{i,j}$ between node j and node i is to 1, the higher the similarity between the two nodes. The closer it is to 0, the lower the similarity between the two nodes. By analogy, the similarity $X_{i,j}$ between all nodes and the target node was calculated, and the nearest $K - 1$ nodes were selected.

The last step was to construct the adjacency matrix A . The i th row and j th column of the adjacency matrix A were the $X_{i,j}$ of node i and node j , respectively. Similarly, the j th row and i th column of the adjacency matrix A were also the $X_{i,j}$ of the nodes i and j , respectively. Because the similarity between the node and itself was 1, the diagonals of

adjacency matrix A were all 1. Thus, we obtained a symmetric matrix with a diagonal of 1, namely adjacency matrix A .

The undirected attributed graph by using the spatio-temporal features and multi-modal features of cellular traffic was constructed successfully, and G was defined as $G = (V, E, S)$, where $V = \{v_1, v_2, \dots, v_K\}$ is the node set. $E = \{e_{12}, e_{23}, e_{34}, \dots\}$ represents the set of undirected edges, corresponding to whether each region is adjacent. $S = \{s_1, s_2, \dots, s_K\}$ represents the the node attribute set. The graph G adeptly integrates the spatio-temporal and geographical attributes of cellular traffic, effectively illustrating the correlation between cellular traffic nodes.

3.2. Dual Branch Spatio-Temporal Graph Neural Network with an Attention Mechanism

In this section, we provide a detailed description of the architecture of DBSTGNN-Att, which was named dual branch spatio-temporal graph neural network with an attention mechanism. DBSTGNN-Att comprised multiple ST-Blocks (two ST-Blocks were used in this paper). Each ST-Block contained two temporal blocks and a spatial block. Matrix P was fed into the temporal block, while the attributed KNN graph was fed into the spatial block. This setup allowed for the extraction of spatio-temporal features from the cellular network traffic. The prediction results Q were ultimately generated through the projection layer. Figure 3 illustrates the structure of DBSGTNN-Att, where the parameter L represents the historical time-series length. The dimension of the matrix S obtained by splicing the multi-modal feature matrix with the attributed KNN matrix was $(L + 1) \times K$, where K is the number of similar nodes extracted and L is the length of the historical time series.

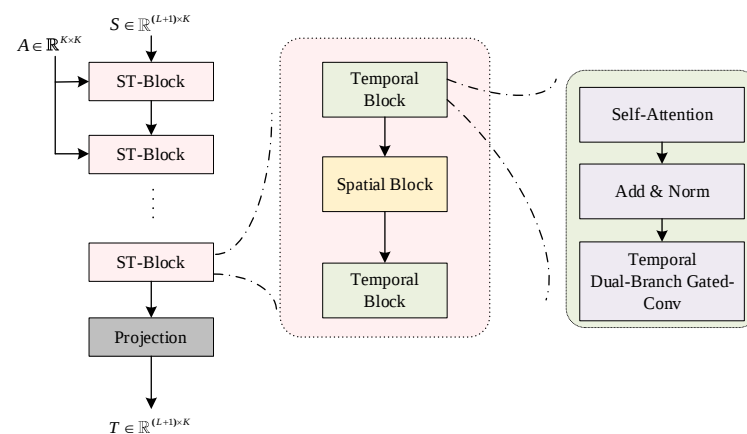


Figure 3. Structure diagram of the DBSTGNN-Att model.

3.2.1. Time Feature Extraction Based on Temporal Dual Branch-Gated Convolution with an Attention Mechanism

The self-attention mechanism is one of the attention mechanisms. The weight assigned to each input item depends on the interaction between the input items, that is, the voting within the input item determines which input items should be focused on. Compared with the strong attention mechanism and the soft attention mechanism, the self-attention mechanism has the advantage of parallel computing when dealing with very long inputs and neutralizes the degree of information neglect and the running time of the strong attention mechanism and the soft attention mechanism. Specifically, the self-attention mechanism captures the dependencies between sequences by the query matrix Q , the key matrix K , and the value matrix V . By comparing the query matrix Q of the time series with the key matrix K , and matching the value matrix V with the value corresponding to the key matrix K , the self-attention can be calculated. By normalizing the time dependence of each time step, the *SoftMax* function can obtain the attention probability matrix. In order to prevent significant differences in the probability distribution of the matrix obtained by the *SoftMax* function after the dot product operation, it is necessary to apply the $\sqrt{d}$ scaling

factor for normalization. Finally, the output is processed by the linear layer and the *ReLU* activation function to obtain the representation of the time feature.

In this paper, we used the self-attention mechanism. Before the temporal dual branch-gated convolution layer, several self-attention layers were augmented to learn feature weights adaptively. One input can generate three matrices: Q (query), K (key), and V (value). The self-attention mechanism is shown in Equations (2) and (3).

$$Attention(Q, K, V) = softmax\left(QK^T / \sqrt{d_k}\right)V \quad (2)$$

$$\begin{cases} Q = W_q S \\ K = W_k S \\ V = W_v S \end{cases} \quad (3)$$

where d_k is the vector dimension of K . S is the time series of the input. W_q , W_k , and W_v are the weight matrices for calculating the query, key, and value, respectively. The self-attention mechanism structure diagram is shown in Figure 4.

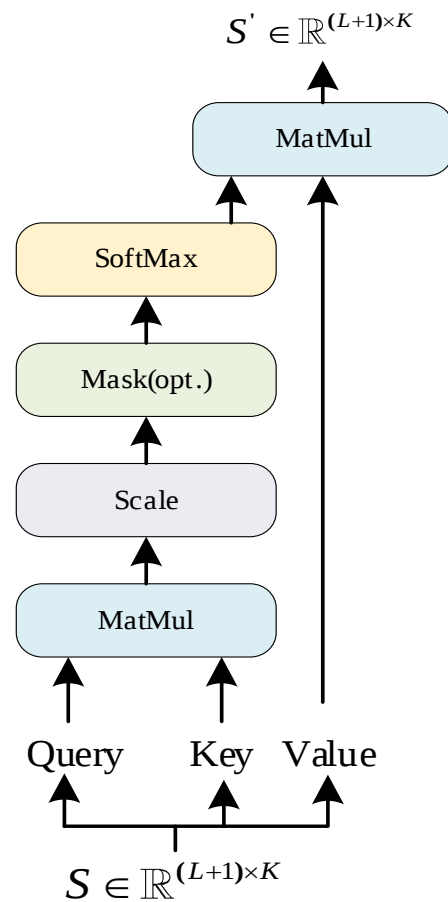


Figure 4. Structure diagram of the self-attention mechanism.

Despite the prevalent usage of RNN and its derivative LSTM for time-series prediction, the computational complexity associated with this type of recurrent neural network infrastructure is considerably high. Conversely, CNN possesses several advantages, such as rapid training speed, uncomplicated structure, and the absence of dependency constraints. Consequently, we introduced the temporal dual branch-gated convolution, augmented with a self-attention mechanism, aiming to effectively discern the temporal features that aid in capturing cellular network traffic. The depiction of the temporal dual branch-gated convolution with a self-attention mechanism structure is illustrated in Figure 5.

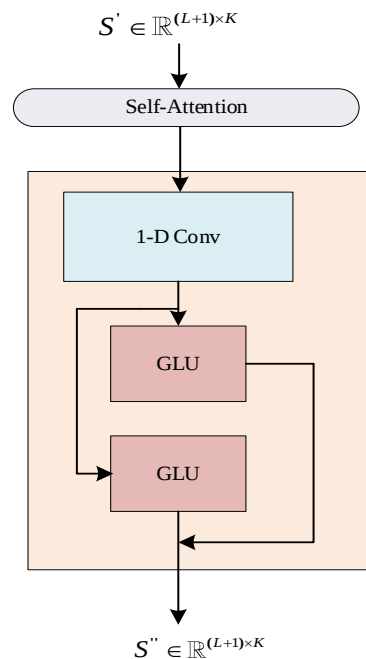


Figure 5. Structure diagram of the temporal block.

Moreover, the 1-D Conv layer divides the matrix into two segments on average, namely S_1 and S_2 . Each segment then separately undergoes processing through two GLUs (gated linear units), after which their respective outputs are combined. The output of the two GLUs is fused to obtain $S'' \in \mathbb{R}^{K \times (L+1)}$. This approach offers an advantage by enabling different GLUs to concentrate on learning distinctive feature representations of various components, thereby augmenting the model's ability to adapt to disparate data models. In the context of predicting cellular network traffic, the traffic patterns may potentially vary due to temporal, geographical, or other contextual factors. Therefore, the adaptability of the model becomes pivotal. The related equations are shown in Equations (4)–(8).

Firstly, the gating ratios g_1 and g_2 were calculated, and the sigmoid activation function was used:

$$g_1 = \sigma(W_{g1} * S_1 + b_{g1}) \quad (4)$$

$$g_2 = \sigma(W_{g2} * S_2 + b_{g2}) \quad (5)$$

where σ represents the sigmoid function, and W and b represent the weight and bias, respectively.

Then, we calculated the inputs h_1 and h_2 of the activation function.

$$h_1 = W_{h1} * S_1 + b_{h1} \quad (6)$$

$$h_2 = W_{h2} * S_2 + b_{h2} \quad (7)$$

Finally, the two outputs were fused to obtain $P' \in \mathbb{R}^{N \times L}$.

$$S'' = g_1 \odot h_1 + g_2 \odot h_2 \quad (8)$$

where $\odot$ represents the Hadamard product and $*$ denotes the convolution operation.

3.2.2. Spatial Feature Extraction Based on Spatial Graph Convolution

The extraction of an intricate spatial correlation is a prominent concern in cellular network traffic prediction. While a traditional convolutional neural network (CNN) can acquire local spatial characteristics, it is restricted to the Euclidean space. Due to the non-two-dimensional grid format, the CNN model fails to depict the complex topology of cellular traffic interchange between regions, resulting in an inadequate capture of spatial interdependence. Lately, the transition of CNN to a graph convolutional network (GCN),

which is capable of handling arbitrary structured data, has garnered considerable attention. The GCN model has proven successful in various applications, including document classification, unsupervised learning, and image classification [31,32].

In this paper, we called the GCN that dealt with spatial features as spatial graph convolution (SGC). Specifically, SGC aggregates the neighbor information and similarity of nodes by performing convolution operations on the graph. At each time step, SGC convolves the features of each node with the features of its neighbor nodes to obtain a new feature representation of the node. In this way, SGC can use the connection relationship between the nodes to capture the spatial dependence in cellular networks. In the DBSTGNN-Attn model, the input of SGC is a graph representing cellular network traffic, where the nodes represent the cellular traffic data detected by detectors in a region, the edges represent the connection relationship between regions, and the weights on the edges represent the similarity between the regions. By convoluting the graph, SGC can aggregate the features of each node with the features of its neighbor nodes to obtain a new feature representation of the node. The specific formula is shown in Equation (9).

$$f(S, A) = \sigma(\hat{A}ReLU(\hat{A}XW_0)W_1) \quad (9)$$

where S represents the attributed matrix, A represents the adjacency matrix, $\hat{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ denotes preprocessing step, $\tilde{A} = A + I_N$ is a matrix with a self-connection structure, and $\tilde{D}$ is the degree matrix, $\tilde{D} = \sum_n \tilde{A}_{nn}$. W_1 and W_0 represent the weight matrix of each layer. $\sigma(\cdot)$ and $ReLU(\cdot)$ represent the activation functions.

4. Results and Discussion

In this section, extensive experimentation is conducted to validate the efficacy of the proposed methodology. Primarily, the datasets, experimental configurations, baseline techniques, and evaluation metrics employed in this study are introduced. To substantiate the superior performance of the proposed approach, a comparative analysis is carried out against the existing methodologies. Ultimately, the experimental findings and performance analysis are presented.

4.1. Datasets Description

In this paper, three datasets were used, among which the cellular traffic dataset used the Milan City cellular traffic dataset provided by Telecom Italia. It was used as part of the “Big Data Challenge” [29]. The dataset collected cellular traffic data for Milan City from 1 November 2013 to 1 January 2014, at a 10 min interval, and collected approximately 19 GB (62 days, 300 million records) of data. In this dataset, the city of Milan is divided into a grid-covered layer of 100×100 squares, each of which is approximately 235×235 m in size, which we called a cell. In each cell, three types of cellular traffic data are included, namely Internet, Call, and SMS. The dataset used in this paper was Internet.

In addition, this paper also used two geographic datasets of Milan City. The dataset on BS information came from OpenCellID [33], which contains multiple types of information about BS, such as location (latitude and longitude), mobile country code, and the coverage of each BS. Using the geographical location information of each cell, we can map the location of each BS to the cell where the BS is located by simple preprocessing. The dataset was not only used in cellular traffic prediction, but also played a key role in the subsequent base station sleep control experiments.

The POI information came from Google Places API [34]. POI refers to the point data in the Internet electronic map, which basically includes four attributes: name, address, coordinate, and category. In the dataset used in this paper, 13 kinds of POIs were collected, including subway stations, shops, restaurants, and parks.

The SA of a region reflects the overall user demand degree for network services. The dataset on social activity level was obtained through Dandelion API [35]. The obtained

data contained the information a user generated when using Twitter, such as the location and keywords.

4.2. Evaluation Indicators

There are many evaluation indexes to evaluate the forecasting effect of the model [36–40], such as the mean absolute error (MAE), root-mean-square error (RMSE), mean absolute percentage error (MAPE), mean squared error (MSE), and R-squared (R²). In this paper, the three most widely used evaluation indicators MAE, RMSE and R² were used. Specifically, MAE is used to measure the gap between the predicted value and the real value. For any size difference, the penalty of MAE is fixed. Therefore, no matter what kind of input value is used, there is a stable gradient, which does not lead to the gradient explosion problem and has a more robust solution. RMSE is the square root of the sum of squares, which enlarges the gap between the larger errors; so, RMSE is more sensitive to errors. Among them, the range of MAE and RMSE is $[0, +\infty]$. When the predicted value is completely consistent with the real value, it is equal to 0, that is, the perfect model. The greater the error, the greater the value. The range of R² is $[0, 1]$. When the predicted value is completely consistent with the real value, it is equal to 1. Their formulas are shown in Equations (10)–(12).

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{Y})^2} \quad (12)$$

Specifically, N is the time steps, $\hat{y}_i$ is the set of prediction values, y is the set of target values, and $\bar{Y}$ is the average of the true observed values.

4.3. Experimental Environment

The experimental environment is the basis of system testing. For the testing of a system, it is necessary to ensure that the environment is configured correctly. The correct experimental environment is a key step in the experiment. In this experiment, we used a deep learning server to configure the experimental environment. The production type of the central processing unit (CPU) was Intel (R) Core (TM) i7-9700, the production type of the GPU was NVIDIA GeForce GTX 745, and the memory size was 16 GB. In addition, Pytorch was used to build a network framework, Adam optimizer was used to iteratively train the same training set partitioning and data preprocessing methods, and Python was used as a programming environment. The experimental environment is shown in Table 1.

Table 1. Experimental environment.

Name	Configuration
Operating System	Windows10
Memory Capacity	16 GB
Hard Disk Capacity	1 TB
CPU	Intel(R) Core(TM) i7-9700
GPU	NVIDIA GeForce GTX 745
Pytorch Version	1.11
Python Version	3.9

4.4. Parameters Settings

In this experiment, we chose a 10×10 cell range in the centre of Milan city and extracted continuous 30-day cellular traffic (Internet) as time series for evaluation. The

cellular traffic data from the first 24 days were used as the training set and the remaining 6 days were used as the test set. Additionally, we set the learning rate of the model to 0.001, the batch size to 32, and the training period to 1500. And Adam was selected as the optimizer. The value of K was set to 10, and K was the number of the nodes of KNN graph. The weights of the POIs, BSs, and SAs (ω_0 , ω_1 , and ω_2) were set to 0.4, 0.4, and 0.2, respectively. The output dimension of the input layer was set to 32, and the (3,1) kernel was used for temporal convolution based on a previous work. In order to prevent overfitting, if the verification loss was not reduced within ten epochs, the model stopped training. The specific parameter settings are shown in Table 2.

Table 2. Model parameter settings.

Model Component	Parameter
Learning rate	0.001
Batch size	32
Training epoch	1500
Optimizer	Adam
K	10
$\omega_0, \omega_1, \omega_2$	0.4, 0.4, 0.2
Output dimension of the input layer	32

4.5. Baseline Models

In order to prove that the cellular traffic prediction method proposed in this paper met the requirements of the base station sleep control strategy in terms of accuracy, we chose statistical models, GCN-based deep learning model variants, and transformer variants as baselines for the comparison.

Autoregressive integrated moving average (ARIMA): ARIMA is a classical time-series analysis and prediction method. It includes three parts, i.e., autoregressive, integrated, and moving averages.

Support vector regression (SVR): SVR is a regression method based on a support vector machine (SVM), which is used to solve regression problems. Unlike traditional regression methods, SVR focuses on finding a function that minimizes the difference between the predicted value and the true value, and tolerates some errors.

Gated recurrent unit (GRU): GRU is a recurrent neural network (RNN) structure that is frequently utilized for the analysis of sequence data. Its primary advantage lies in its ability to effectively capture long-term dependencies within the sequence through the introduction of a gating mechanism. Consequently, this enhances the overall performance of the model.

Temporal graph convolutional network (T-GCN): T-GCN is an artificial neural network model that integrates the graph convolutional network (GCN) and gated recurrent unit (GRU). Its purpose is to effectively capture the spatio-temporal correlation within traffic data and thereby achieve accurate cellular network traffic prediction.

Spatial-temporal self-attention graph convolutional network (STA-GCN): STA-GCN is a spatial-temporal self-attention graph convolutional network with a good prediction performance and interpretability.

4.6. Experimental Results

The evaluation results comprise numerical comparisons using MAE, RMSE, and R^2 , as well as performance displays, between DBSTGNN-Att and the five other prediction models. These comparisons were based on the cellular traffic dataset of Milan. The prediction was compared in three different scenarios, namely short-term prediction, mid-term prediction, and long-term predictions, with sampling intervals of 10 min, 30 min, and 60 min, respectively. When the sampling interval was 10 min, the model predicted the cellular traffic in the next one day; when the sampling interval was 30 min, the model predicted the cellular traffic in the next three days; and when the sampling interval was

60 min, the model predicted the cellular traffic in the next six days. In order to demonstrate the prediction performance of DBSTGNN-Att across various sampling intervals of cellular traffic, a representative cell (ID: 4259), which contains the most famous university in Milan named Bocconi, was selected to exhibit the prediction results of the test dataset. Figure 6 provides an intuitive representation of the prediction comparisons for the last 144 time points, representing short-term (one day) and mid-term (three days) predictions, in the selected cell. Furthermore, the numerical comparisons for the aforementioned prediction performance are presented in Table 3 (Note: The results shown in this section do not use the attributed KNN graph to fuse multi-modal features).

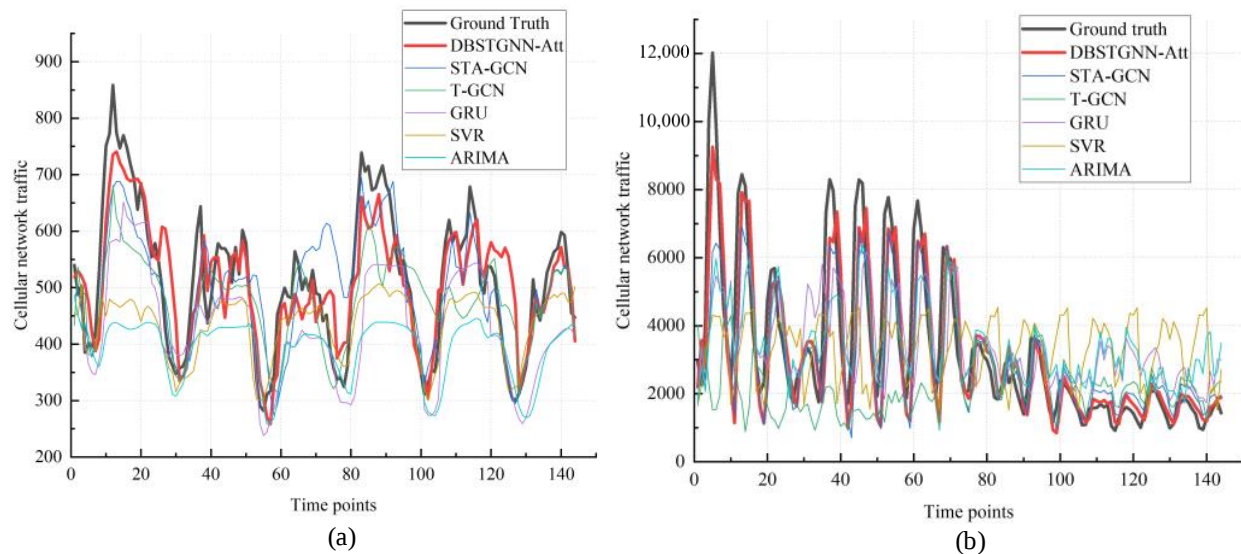


Figure 6. Prediction performance display of cell No. 4259 for different sampling intervals, with each time point representing a sampling interval, for a total of 144 time points. (a) Short-term prediction: 10-min sampling interval; and (b) mid-term prediction: 30-min sampling interval.

Table 3. Numerical comparisons of the prediction performance.

Sampling Interval	10 min			30 min			60 min		
Model Metrics	MAE	RMSE	R ²	MAE	RMSE	R ²	MAE	RMSE	R ²
ARIMA	58.24	102.33	0.53	178.44	216.23	0.48	378.55	453.11	0.27
SVR	52.67	85.31	0.57	176.24	211.65	0.50	340.59	407.26	0.30
GRU	34.22	41.72	0.77	104.51	129.27	0.63	235.11	276.50	0.41
T-GCN	27.96	40.78	0.81	79.22	122.40	0.69	170.16	232.58	0.50
STA-GCN	26.58	38.71	0.84	75.58	115.43	0.66	157.96	211.12	0.58
DBSTGNN-Att	26.02	37.98	0.83	71.24	106.43	0.71	149.81	197.13	0.65

In the short-term prediction scenario, compared to the GRU, T-GCN, and STA-GCN, DBSTGNN-Att presents performance improvements of around 23.96%, 6.94%, and 2.11% in the MAE and 8.96%, 6.87%, and 1.89% in the RMSE, respectively. In the mid-term prediction scenario, compared to the GRU, T-GCN, and STA-GCN, DBSTGNN-Att presents performance improvements of around 31.83%, 10.07%, and 5.74% in the MAE and 17.67%, 13.05%, and 7.80% in the RMSE, respectively. In the long-term prediction scenario, compared to the GRU, T-GCN, and STA-GCN, DBSTGNN-Att presents performance improvements of around 36.28%, 11.94%, and 5.16% in the MAE and 28.71%, 15.24%, and 6.63% in the RMSE, respectively.

This is due to the ARIMA and SVR having difficulties in dealing with complex non-linear cellular traffic changes. Specifically, the ARIMA is a classical time-series analysis and prediction algorithm, which has a good effect on simple linear trend and periodic

change data. However, cellular traffic data often have complex nonlinear characteristics, which makes it difficult for ARIMA to accurately capture these complex change patterns, especially in mid-term and long-term predictions. The SVR is a machine learning algorithm that can capture nonlinear time features. Nevertheless, in cellular traffic prediction, cellular traffic data usually have a high degree of noise and uncertainty, which makes it difficult for the SVR to accurately model and predict. Secondly, the SVR needs to manually select kernel functions and adjust hyperparameters, which is a challenge for the complexity of cellular traffic data. As a deep learning algorithm, the GRU has produced good results in time-series prediction. But the GRU only processes the time characteristics of cellular network traffic data and does not consider the spatial distribution of cellular traffic; so, the GRU is more suitable for single-region cellular traffic prediction without spatial characteristics. In this scenario, its simple structure and faster running speed are its advantages. In contrast, T-GCN and STA-GCN not only consider the time characteristics of cellular traffic, but also use GCN to process the spatial characteristics of cellular traffic, which greatly improves the prediction accuracy. Nevertheless, T-GCN does not use the attention mechanism module, so that it may repeatedly learn similar spatial features, resulting in a low training efficiency and model performance degradation. STA-GCN only uses the structure of one GLU, which causes the GLU to learn the feature representation of multiple different components at the same time, which reduces the adaptability of the model.

The efficacy of DBGSTNN-Att in predicting cellular network traffic was substantiated through numerical comparisons to and the analysis of several models. This is attributed to the attention mechanism of DBGSTNN-Att, which prioritizes the periodic changes in cellular traffic. Furthermore, the use of the dual GLU module enables distinct GLUs to concentrate on learning the distinctive feature representation of various components. As a result, the model's adaptability to different data is enhanced, significantly boosting the accuracy of the prediction.

4.7. Ablation Study

To further investigate the effectiveness of the modules in DBSTGNN-Att, we conducted evaluations for three DBSTGNN-Att variants: (1) W/o an attributed KNN graph: It does not use the attributed KNN graph as a model input and does not fuse multi-model features (BSs and POIs). (2) W/o a dual GLU: The model does not use a double GLU structure. (3) W/o an attention mechanism: Dual branch-gated convolution does not use the attention mechanism. The prediction performance comparisons of the ablation study of DNSTGNN-Att are shown in Table 4, where w/o denotes the deletion of a component.

By examining Table 4, it is evident that the aforementioned three components have a positive impact on the prediction of cellular traffic. Among these components, the dual GLU structure exhibits the greatest improvement in the model's prediction accuracy. This can be attributed to the fact that the dual GLU structure enhances the model's flexibility, enabling it to better adapt to diverse data distributions and complex models. Additionally, each GLU has the capacity to learn distinct feature representations, which are then combined through element-level multiplication operations. This integration allows the model to more effectively capture the nonlinear relationships present in the input data, thereby enhancing the model's expressive capabilities and prediction performance.

Furthermore, the attributed KNN graph incorporates BSs, POIs, and SAs, encompassing a wider range of information that is beneficial for cellular traffic prediction. The multi-modal features are fused and input into the neural network, which increases the interpretability of the model and improves the prediction accuracy.

Table 4. Prediction performance comparisons of the ablation study of DNSTGNN-Att.

Sampling Interval	10 min		30 min		60 min	
Model Metrics	MAE	RMSE	MAE	RMSE	MAE	RMSE
DBSTGNN-Att	24.23	35.22	68.01	95.46	137.11	179.57
W/o the attributed KNN graph	26.02	37.98	71.24	106.43	149.81	197.13
W/o a dual GLU	27.66	43.83	78.45	123.14	159.35	218.70
W/o an attention mechanism	25.39	35.71	80.72	129.94	172.36	257.97

The influence of the attention mechanism on the prediction accuracy remains minimal in short-term prediction. However, as the collection interval of cellular traffic data expands, the advantageous impact of the attention mechanism on the prediction accuracy becomes increasingly prominent. This is because the attention mechanism enables the model to assign different weights to different parts of the input, thereby paying more attention to information that is more important to the current task. This mechanism enables the model to better handle long sequences and complex inputs, and improves the expression ability and prediction performance of the model.

4.8. The Influence of the Hidden Unit and Sliding Window Lengths on the Experimental Results

The setting of hidden units and sliding windows has a greater impact on the prediction results of the model. In the choice of the two, the hidden unit is too large, which may lead to an increased complexity of the model and an over-fitting phenomenon. The larger the sliding unit, the more features are perceived, which may slow down the running speed of the model and reduce the prediction accuracy. Therefore, we tested the settings of the hidden unit and the sliding window. The range of the number of hidden units was set from 32 to 128, and the range of the sliding window was set from 3 to 6. It can be seen from Figures 7 and 8 that, in the case of three sampling intervals, when the number of hidden units is 96, the prediction effect of the model is the best. In the case of three sampling intervals, when the sliding window is 9, the prediction effect of the model is the best. The performance comparison under different hidden units is shown in Figure 7, where the x -axis represents the number of hidden units and the y -axis represents the values of the MAE and the RMSE. The performance comparison under different sliding window lengths is shown in Figure 8, where the x -axis represents the length of the sliding window and the y -axis represents the values of the MAE and the RMSE.

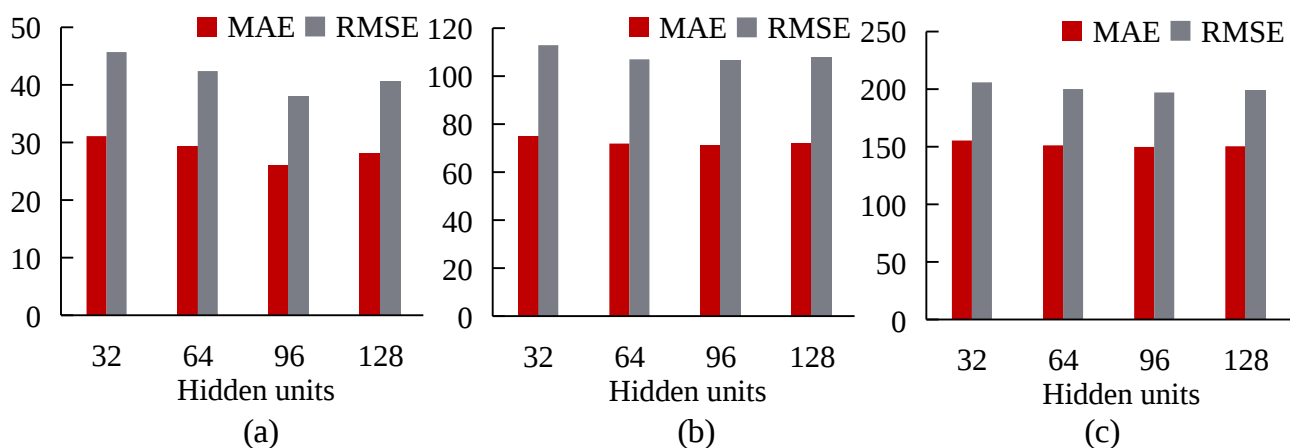


Figure 7. Performance comparison under different hidden units. (a) A 10 min sampling interval; (b) a 30 min sampling interval; and (c) a 60 min sampling interval.

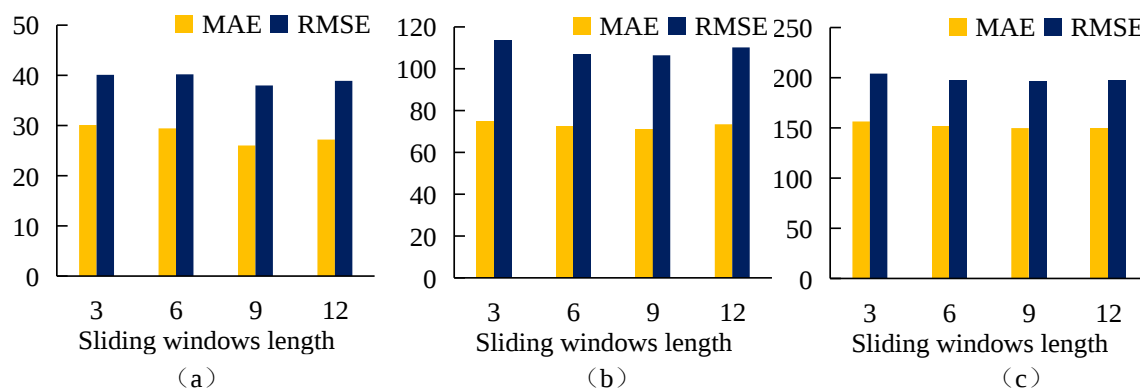


Figure 8. Performance comparison under different sliding window lengths. (a) A 10 min sampling interval; (b) a 30 min sampling interval; and (c) a 60 min sampling interval.

5. Limitations and Prospects

There are some limitations to this study, which are caused by the complexity of calculations and the difficulty of data acquisition. Specifically, when selecting the most similar K nodes, it is necessary to calculate the similarity between the target node and all the remaining nodes, and select the $(K - 1)$ nodes with the highest similarity to generate the KNN graph with the target node. This process requires a large quantity of calculations, which increases the running time of the entire model. Secondly, in order to be able to predict the random fluctuations in cellular network traffic, more multi-modal data should be considered, such as weather, traffic conditions, and holiday information. However, due to the large differences in data distribution between different countries and regions, it is necessary to use data according to the specific situation in practice.

In the future, we can consider using clustering methods (such as K-means and mutual information) to cluster nodes and generate a new graph for nodes in each cluster to input into the model for prediction to reduce the complexity of calculations. In addition, with the gradual popularization of artificial intelligence models, models currently used in popular fields, such as natural language processing, can be applied to the field of cellular traffic prediction. For example, the proposed TimeGPT model is based on Transformer. Therefore, it is a meaningful work to combine the spatio-temporal convolution used in this paper with the large model. In addition, from the perspective of exploring multi-modal features, more features that contribute to cell flow prediction can be combined to improve prediction accuracy. This is because multi-modal features can provide a richer information representation for the model.

6. Conclusions

In this paper, we commenced by presenting the pertinent background in the area of cellular traffic prediction. Subsequently, we reviewed recent research, ranging from the conventional regression model to various iterations based on graph neural networks. By succinctly summarizing the challenges and drawbacks of the existing research, we proposed a novel approach titled dual branch spatio-temporal graph neural network with an attention mechanism for cellular network traffic prediction. In addition, this study also used the attributed KNN graph. By rearranging the K most similar nodes and fusing multi-modal features, the accuracy of cellular traffic prediction was improved. To substantiate the efficacy and enhancement in the performance of DBSTGNN-Att, we initially compared it with the existing models, such as the ARIMA, the GRU, and T-GCN, using a real-world dataset, subsequently analyzing the experimental outcomes. The results demonstrate the superiority of the proposed method over the aforementioned compared models in short-term, mid-term, and long-term prediction scenarios.

Author Contributions: Conceptualization, Z.C. and C.T.; methodology, Z.C. and C.T.; validation, J.Z., L.Z. and Y.F.; data curation, Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (62072416), the Key Research and Development Special Project of Henan Province (221111210500), and the Key Technologies R&D Program of Henan Province (232102211053, 222102210170, and 222102210322).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Gupta, A.; Jha, R.K. A survey of 5G network: Architecture and emerging technologies. *IEEE Access* **2015**, *3*, 1206–1232. [\[CrossRef\]](#)
- Khanh, Q.V.; Hoa, N.V.; Manh, L.D.; Le, A.N.; Jeon, G. Wireless communication technologies for IoT in 5G: Vision, applications, and challenges. *WCMC* **2022**, *2022*, 3229294. [\[CrossRef\]](#)
- Izadi, S.; Ahmadi, M.; Rajabzadeh, A. Network traffic classification using deep learning networks and Bayesian data fusion. *J. Netw. Syst. Manag.* **2022**, *30*, 5. [\[CrossRef\]](#)
- Zhao, J.; Jing, X.; Yan, Z.; Pedrycz, W. Network traffic classification for data fusion: A survey. *Inf. Fusion* **2021**, *30*, 22–47. [\[CrossRef\]](#)
- Javaheri, D.; Gorgin, S.; Lee, J.-A.; Masdari, M. Fuzzy logic-based DDoS attacks and network traffic anomaly detection methods: Classification, overview, and future perspectives. *Inf. Sci.* **2023**, *626*, 315–338. [\[CrossRef\]](#)
- Liu, X.; Zhang, Z.; Hao, Y.; Zhao, H.; Yang, Y. Optimized OTSU Segmentation Algorithm-Based Temperature Feature Ex-traction Method for Infrared Images of Electrical Equipment. *Sensors* **2024**, *24*, 1126. [\[CrossRef\]](#)
- Meng, T.; Jing, X.; Yan, Z.; Pedrycz, W. A survey on machine learning for data fusion. *Inf. Fusion* **2020**, *57*, 115–129. [\[CrossRef\]](#)
- Zhang, C.; Zhang, H.; Qiao, J.; Yuan, D.; Zhang, M. Deep Transfer Learning for Intelligent Cellular Traffic Prediction Based on Cross-Domain Big Data. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1389–1401. [\[CrossRef\]](#)
- Shahi, T.B.; Shrestha, A.; Neupane, A.; Guo, W. Stock Price Forecasting with Deep Learning: A Comparative Study. *Mathematics* **2020**, *8*, 1441. [\[CrossRef\]](#)
- Yun, K.K.; Yoon, S.W.; Won, D. Interpretable stock price forecasting model using genetic algorithm-machine learning re-gressions and best feature subset selection. *Expert Syst. Appl.* **2023**, *213*, 118803. [\[CrossRef\]](#)
- Braz, F.J.; Ferreira, J.; Gonçalves, F.; Weege, K.; Almeida, J.; Baldo, F.; Gonçalves, P. Road traffic forecast based on meteorological information through deep learning methods. *Sensors* **2022**, *22*, 4485. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wang, Y.; Fang, S.; Zhang, C.; Xiang, S.; Pan, C. TVGCN: Time-variant graph convolutional network for traffic forecasting. *Neurocomputing* **2022**, *471*, 118–129. [\[CrossRef\]](#)
- Chuwang, D.D.; Chen, W.; Zhong, M. Short-term urban rail transit passenger flow forecasting based on fusion model methods using univariate time series. *Appl. Soft Comput.* **2023**, *147*, 110740. [\[CrossRef\]](#)
- Li, H.; Jin, K.; Sun, S.; Jia, X.; Li, Y. Metro passenger flow forecasting through multi-source time-series fusion: An ensemble deep learning approach. *Appl. Soft. Comput.* **2022**, *120*, 108644. [\[CrossRef\]](#)
- Choi, B. *ARMA Model Identification*; Springer Science & Business Media: Berlin, German, 2012; pp. 1–27.
- Dey, B.; Roy, B.; Datta, S.; Ustun, T.S. Forecasting ethanol demand in India to meet future blending targets: A comparison of ARIMA and various regression models. *Energy Rep.* **2023**, *9*, 411–418. [\[CrossRef\]](#)
- Azari, A.; Papapetrou, P.; Denic, S.; Peters, G. Cellular Traffic Prediction and Classification: A Comparative Evaluation of LSTM and ARIMA. In Proceedings of the Discovery Science: 22nd International Conference, Split, Croatia, 28–30 October 2019; pp. 129–144. [\[CrossRef\]](#)
- Yang, H.; Li, X.; Qiang, W.; Zhao, Y.; Zhang, W.; Tang, C. A network traffic forecasting method based on SA optimized ARIMA-BP neural network. *Comput. Netw.* **2021**, *193*, 108102. [\[CrossRef\]](#)
- Chauhan, V.K.; Dahiya, K.; Sharma, A. Problem formulations and solvers in linear SVM: A review. *Artif. Intell. Rev.* **2018**, *52*, 803–855. [\[CrossRef\]](#)
- Zhang, K.; Chuai, G.; Zhang, J.; Chen, X.; Si, Z.; Maimaiti, S. DIC-ST: A Hybrid Prediction Framework Based on Causal Structure Learning for Cellular Traffic and Its Application in Urban Computing. *Remote Sens.* **2022**, *14*, 1439. [\[CrossRef\]](#)
- Lai, J.; Chen, Z.; Zhu, J.; Ma, W.; Gan, L.; Xie, S.; Li, G. Deep Learning Based Traffic Prediction Method for Digital Twin Network. *Cogn. Comput.* **2023**, *15*, 1748–1766. [\[CrossRef\]](#)
- Yang, L.; Gu, X.; Shi, H. A Novel Satellite Network Traffic Prediction Method Based on GCN-GRU. In Proceedings of the 2020 International Conference on Wireless Communications and Signal Processing, Nanjing, China, 21–23 October 2020; pp. 718–723. [\[CrossRef\]](#)

23. Zhang, S. A prediction model of Network traffic noise reduction based on PSO-VMD and BiLSTM. In Proceedings of the 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications, Changchun, China, 20–22 May 2022; pp. 1–7. [\[CrossRef\]](#)
24. Peng, D.; Zhang, Y. MA-GCN: A Memory Augmented Graph Convolutional Network for traffic prediction. *Eng. Appl. Artif. Intell.* **2023**, *121*, 106046. [\[CrossRef\]](#)
25. Kumar, R.; Mendes; Moreira, J.; Chandra, J. DyGCN-LSTM: A dynamic GCN-LSTM based encoder-decoder framework for multistep traffic prediction. *Appl. Intell.* **2023**, *53*, 1–24. [\[CrossRef\]](#)
26. Shen, W.; Zhang, H.; Guo, S.; Zhang, C. Time-Wise Attention Aided Convolutional Neural Network for Data-Driven Cellular Traffic Prediction. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1747–1751. [\[CrossRef\]](#)
27. Wang, Z.; Hu, J.; Min, G.; Zhao, Z.; Chang, Z.; Wang, Z. Spatial-Temporal Cellular Traffic Prediction for 5G and Beyond: A Graph Neural Networks-Based Approach. *IEEE Trans. Ind. Inform.* **2022**, *19*, 5722–5731. [\[CrossRef\]](#)
28. Zhou, X.; Zhang, Y.; Li, Z.; Wang, X.; Zhao, J.; Zhang, Z. Large-scale cellular traffic prediction based on graph convolutional networks with transfer learning. *Neural Comput. Appl.* **2022**, *34*, 5549–5559. [\[CrossRef\]](#)
29. Barlacchi, G.; De Nadai, M.; Larcher, R.; Casella, A.; Chitic, C.; Torrisi, G.; Antonelli, F.; Vespignani, A.; Pentland, A.; Lepri, B. A multi-source dataset of urban life in the city of Milan and the Province of Trentino. *Sci. Data* **2015**, *2*, 150055. [\[CrossRef\]](#)
30. Wang, Z.; Rosen, D. Manufacturing process classification based on heat kernel signature and convolutional neural networks. *J. Intell. Manuf.* **2023**, *34*, 3389–3411. [\[CrossRef\]](#)
31. Xue, Z.; Liu, Z.; Zhang, M. DSR-GCN: Differentiated-Scale Restricted Graph Convolutional Network for Few-Shot Hyper-spectral Image Classification. *IEEE Trans. Geosci. Remote* **2023**, *61*, 1–18. [\[CrossRef\]](#)
32. Cui, H.; Wang, G.; Li, Y.; Welsch, R.E. Self-training method based on GCN for semi-supervised short text classification. *Inf. Sci.* **2022**, *611*, 18–29. [\[CrossRef\]](#)
33. OpenCellID. The World's Largest Open Database of Cell Towers. Available online: <https://opencellid.org/> (accessed on 13 January 2024).
34. Google Inc. Google Places API. Available online: <https://developers.google.com/places/> (accessed on 13 January 2024).
35. Dandelion. Dandelion API. Available online: <https://dandelion.eu> (accessed on 13 January 2024).
36. Nabi, S.T.; Islam, R.; Alam, G.R.; Hassan, M.M.; AlQahtani, S.A.; Aloji, G.; Fortino, G. Deep Learning Based Fusion Model for Multivariate LTE Traffic Forecasting and Optimized Radio Parameter Estimation. *IEEE Access* **2023**, *11*, 14533–14549. [\[CrossRef\]](#)
37. Gu, B.; Zhan, J.; Gong, S.; Liu, W.; Su, Z.; Guizani, M. A Spatial-Temporal Transformer Network for City-Level Cellular Traffic Analysis and forecasting. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 9412–9423. [\[CrossRef\]](#)
38. Xiong, Z.; Zhang, K.; Chuai, G.; Yang, X.; Xu, Y. Intelligent Cellular Traffic Prediction in Open-RAN Based on Cross-Domain Data Fusion. In Proceedings of the IEEE INFOCOM 2023—IEEE Conference on Computer Communications Workshops (Infocom Wkshps), Hoboken, NJ, USA, 20 May 2023; pp. 1–6.
39. Jiang, P.; Zhang, Z.; Dong, Z.; Yang, Y.; Pan, Z.; Yin, F.; Qian, M. Transient-steady state vibration characteristics and influencing factors under no-load closing conditions of converter transformers. *Int. J. Electr. Power Energy Syst.* **2024**, *155*, 109497. [\[CrossRef\]](#)
40. Dan, Y.; Zhang, Z.; Yin, J.; Yang, J.; Deng, J. Parameters estimation of horizontal multilayer soils using a heuristic algorithm. *Electr. Power Syst. Res.* **2022**, *203*, 107661. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.