Spatiotemporal Attention Networks for Traffic Demand Prediction

Guannan Liu[1], and Chenxi Chen[1], and Xin Wan[2]*, and Junjie Wu[1]

[1]*Beihang University, Beijing, China*
[2]*National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing, China*

Abstract

Travel demand, which include taxi, bus, and bike demand, forecasting the travel demand is an important part of intelligent city intelligent transportation system. Accurate prediction models can help cities pre allocate resources to meet travel demand, reduce energy waste. Travel demand prediction can be summarized as spatiotemporal sequence prediction. For a long time, in the field of spatiotemporal sequence prediction, most of them will emphasize the effective capture and modeling of nonlinear and complex spatiotemporal dependency, which can effectively improve the accuracy of spatiotemporal sequence prediction. At the same time, the demand for multi-step prediction is also increasing. Long-time high-precision prediction can effectively improve the auxiliary role of the model for decision-making. To address these issues, we propose a Spatiotemporal Attention Network (STATTN) with a novel spatiotemporal attention mechanism that capture dependency in time-dimension and spatial-dimension at the same time which is spatiotemporal dependency. In order to learn high-quality representation of spatial points in spatiotemporal sequence units, we adopt dilated temporal 1d convolutional neural networks which has ability to learn representation from data through back propagation. To alleviate the error propagation, we use the generate-style decoder which can generate the output without iteration steps. Through extensive experiments on two prediction tasks, we demonstrate the advantages of STATTN in short-term and long-term prediction scenarios.

Keywords

Traffic prediction, spatiotemporal prediction, self-attention network, deep learning

1. Introduction

Nowadays, with the increasing popularity of taxi service platforms such as Uber, LYFT and Didi, people are more willing to call a car on their smartphone than take a taxi at will. In most urban areas, the supply and demand of online car hailing services is unbalanced. For passengers, there is a situation that there are no taxis nearby in some periods of time. Meanwhile, some taxi drivers spend too much time roaming empty cars in other areas. Ultimately, these areas are divided into oversupply and oversupply. On the one hand, this leads to the loss of profits of drivers and taxi companies. On the other hand, it is a waste of time for passengers.

Traffic prediction is one of the most basic problems in intelligent transportation system. In particular, travel demand forecasting is very important for traditional taxi services and online hailing systems (such as Uber, LYFT and Didi travel). In recent years, with mobile devices and wireless communication technology making a large amount of traffic data easy to obtain, travel demand forecasting has become an increasingly promising tool to balance vehicle supply and demand with low-cost and high-quality services, which will create greater economic profits.

ICBASE2022@3rd International Conference on Big Data & Artificial Intelligence & Software Engineering, October 21-23, 2022, Guangzhou, China
wanxin@cert.org.cn (Xin Wan)
© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3304/paper16.pdf

With the fast development of artificial intelligence technology, the newly developed CNN, GCN and GAT network structures can effectively support the processing of spatiotemporal sequence unit data. While the sequence data in the time dimension we could use RNN, LSTM or Transformer which is brilliant in the field of natural language processing, and its variant informer model in the field of time series prediction can provide important support for our prediction work.

Recent years' research progress in spatiotemporal sequence modeling and prediction have proved that the unique spatiotemporal correlation modeling method can effectively improve the accuracy of model prediction. Take the peak demand of taxis as an example, which usually occurs in the business district, urban concentrated residential areas may have peak demand during working hours at the same time. These are important temporal and spatial relationships, so it is quite necessary to model these relationships.

To cope with aforementioned issues, we proposed a model called STATTN. Specifically, we use Dilated TCN to generate the representation of each spatial node and dedicate spatiotemporal self-attention network to capture spatiotemporal dependencies efficiently, carefully designed encoder and decoder make contribution to help us get high quality spatiotemporal representation and avoid error accumulation in inference period.

2. Related work

Traditional spatiotemporal series prediction models are generally based on the transformation of classical time series models, such as ARIMA and STARIMA. Some models are based on feature engineering to construct temporal and spatial relationship features, and use regression model to complete prediction work.

Compared with the traditional time series problem, predicting spatiotemporal series is more challenging, because it deals with not only nonlinear time correlation, but also dynamic and complex spatial correlation [1]; Moreover, whether the traffic flow of urban road network or the travel demand of taxi and bicycle, the spatiotemporal observation of travel is not independent at each location and time point, and there is dynamic correlation. Therefore, the key to solve such problems is to effectively extract the spatiotemporal correlation of data [2], that is, spatiotemporal correlation modeling. Moreover, predicting the long-term future has become one of the most urgent needs of urban computing systems. More and more urban operations need several hours of preparation before the final decision, such as dynamic traffic management and intelligent service allocation [3].

ConvLSTM model changes the one-dimensional vector processed by the traditional LSTM model into a three-dimensional tensor, that is, it adds two dimensions of rows and columns to the original measurement vector. By setting the LSTM learning parameter as tensor, the original matrix multiplication is transformed into convolution operation, and the tensor input processing is realized to complete the task of spatiotemporal sequence prediction [4]. After the release of ConvLSTM model, there is no essential optimization in this structure. The research in the next few years is basically based on the simple expansion of this unit [5], application [6], and multi view training [7]. Until the release of SA ConvLSTM [8] model in 2020, the self attention mechanism is added to the ConvLSTM unit to capture long-range spatiotemporal correlation, which is a great progress.

STGCN model [9], proposed in 2018, uses graph neural network (GCN) to complete the processing of timing units, and creatively proposes the gated time convolution structure to complete the prediction of traffic flow. Subsequently, in 2019, the team of Beijing Jiaotong University released the STGCN based on the attention mechanism version at the AAAI conference, that is, the ASTGCN model [2] to complete the traffic flow prediction. In 2020, the team released the STSGCN model [19] which can extract the temporal and spatial correlation at the same time to complete the traffic flow prediction, and achieved better prediction performance.

Recently, lots of researches begin to focus on model the spatiotemporal correlation of spatiotemporal sequences in their respective deep learning models, so as to achieve better prediction results. Lin's research [1], MSA (multi space attention) mechanism is proposed to model spatiotemporal correlation; In Song's research [10], STGCM (spatial temporal synchronous graph neural module) was proposed to model local spatiotemporal correlation; In order to obtain global spatiotemporal correlation information, in the research of Li [11], the gated CNN module is

assembled in parallel with the spatial temporal fusion graph module. The remote spatiotemporal correlation can be extracted by stacking more module layers; In Fang's research [12], inspired by node [13] and xhoneux's work on continuity graph neural network [14], GCN can be understood as a discrete form of ODE (first-order differential equation). The introduction of ode can effectively avoid the problem of over smoothing caused by the increase of GCN layers and retain effective information for better spatiotemporal correlation modeling. In addition, the use of graph attention [15] is also a highly available method for modeling long-distance spatiotemporal correlation. For example, Fang's research in 2020 [16], using 3DGAT module to model spatiotemporal correlation and complete travel time estimation work.

3. Preliminaries

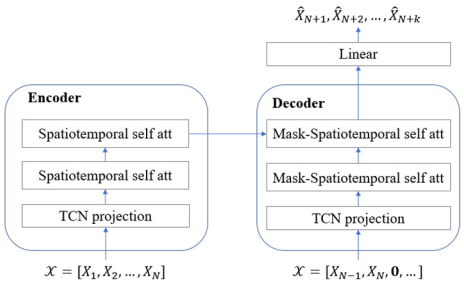

Figure 1. Structure of STATTN

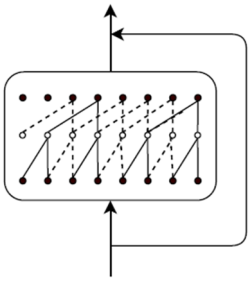

Figure 2. Structure of DTCN

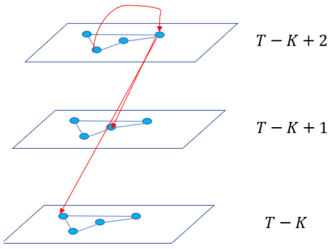

Figure 3. Sketch Map of Spatiotemporal Attention

3.1. Problem Formulation

Given the tensor $\mathcal{X} \in \mathbb{R}^{T \times N \times F}$ observed on a traffic demand map, our goal of traffic demand prediction is to fitting a mapping function f from the historical T observations to predict the future T' traffic demand observations.

$$[X_{t-T+1}, X_{t-T+2}, \ldots, X_t] \rightarrow^f [X_{t+1}, X_{t+2}, \ldots, X_{t+T'}] \tag{1}$$

We denote $X \in \mathbb{R}^{N \times F}$ and $\mathcal{X} = [X_1, X_2, \ldots, X_T]^T$, N denotes the grid numbers or observation numbers in the traffic demand map, F denotes the dimension of observation feature.

3.2. Self-Attention Mechanism

In most existing research on series processing, such as nature language processing, graph representation and time series representations, they[15][17][18]use self-attention mechanism to capture important dependency between neighbors in time axis or graphs in parallel. We define the input as $X \in \mathbb{R}^{N \times F}$, and duplicate input X to generate queries, keys and values, $Q, K, V \in \mathbb{R}^{N \times F}$, and compute output of self-attention mechanism as:

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

3.3. Tensor multiplication

We define tensors $\boldsymbol{\tau} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ and $\boldsymbol{\upsilon} \in \mathbb{R}^{d_1 \times d_2 \times d_4}$, and computing the tensor multiplication as:

$$(\tau \times_2 v)_{ilk} = \sum_{j=1}^{d_2} \tau_{ijk} \cdot v_{ijl} \tag{3}$$

4. Methodology

4.1. Dilated Temporal Convolutional Network

Most researches on spatiotemporal sequence prediction will use the date meaning of time points in time series, such as Monday or Tuesday in a week and which time period of the day to represent the spatial points of time series units. At the same time, many manually set semantic tags will be used to represent the spatial points in time series units. Most of the methods mentioned above require knowledge in relevant fields and many manual judgments. Among them, the parameters of this part cannot be optimized through gradient information, which deviates from the original intention of our data-driven model optimization. Therefore, in the problem of spatial points of sequential units, we use dilated 1-d temporal-axis convolutional network to capture the long-time dependency and use C channels to get the representation of the spatial points in time series units simultaneously.

$$H_{TCN}^{l} = \begin{cases} \mathcal{X} & , l = 0 \\ \sigma\left(W^{l} *_{d^l} H_{TCN}^{l-1}\right), & l = 1,2, \dots, N \end{cases} \tag{4}$$

where $\mathcal{X} \in \mathbb{R}^{T \times N \times F}$is the input of TCN, $H_{TCN}^{l} \in \mathbb{R}^{T \times N \times C}$ is the output of the l-th layer of TCN, andW^{l}denotes the l-th convolution kernel. To expand the receptive field, an exponential dilation rate$d^{l} = 2^{l} - 1$ is adopted in temporal convolution. In the process, zero-padding strategy is utilized to keep time sequence length unchanged.

4.2. Spatiotemporal Self-attention Networks

In order to better obtain the characteristics of spatiotemporal Association of long-term and multi neighbours, so as to encode the historical information, so as to obtain better temporal representation for the subsequent spatiotemporal sequence prediction task, we propose spatiotemporal self-attention network to capture dependency in time-dimension and spatial -dimension at the same time. We organize all history information as $\mathcal{X} \in \mathbb{R}^{T \times N \times F}$, then we use dilated temporal convolutional network(TCN) to generate node representation in spatiotemporal units, which means we could use TCN to get queries, keys and values in parallel:

$$Q, K, V = TCN(\mathcal{X}), TCN(\mathcal{X}), TCN(\mathcal{X}) \in \mathbb{R}^{T \times N \times C} \tag{5}$$

Then, we calculate the attention score tensor as follows:

$$Scores = Q \times_2 K \in \mathbb{R}^{T \times N \times T \times N} \tag{6}$$

However, according to common sense, we know that it is impossible for all spatial points to have an interactive impact. In the decoder, we also need to consider that it is impossible for the later time point to have an impact on the previous time point. Therefore, we need to add mask after completing the calculation of attention score to prevent the above situation, so as to enhance the performance of the model. In the previous experiment, it has been shown that the addition of mask can effectively enhance the performance of the model in the prediction task.

$$Scores_{mask} = Scores \times Mask_{spatial} \times Mask_{temporal} \tag{7}$$

The design and calculation of temporal-mask and spatial-mask will be described in detail in subsequent chapters. After the calculation of attention score is completed and the residual connection is added, the calculation of single-layer spatiotemporal self-attention is completed：

$$X^{l} = X^{l-1} + ReLu\left(softmax\left(\frac{Scores}{\sqrt{E}}\right) \times_4 V\right) \tag{8}$$

4.3. Spatiotemporal Encoder

Before explaining the design of encoder in detail, it is necessary to talk about the design and calculation process of spatial mask. In the previous research work of spatiotemporal sequence prediction, many models will define the adjacency relationship of spatial nodes in advance. Some take the historical sequence similarity of each node as the measurement index, and some take the distance of physical space as the measurement index. However, the above methods have to face the problem of manually determining the threshold to determine whether there is an edge between two points. The manually selected parameters, also non data-driven parameters, will have a certain impact on the performance of the model. Therefore, we use the learnable graph structure to obtain the spatial mask required by the encoder and subsequent decoder.

As we all know, we can decompose the adjacency matrix of the graph as:

$$\boldsymbol{A} = \boldsymbol{E_1 E_2^T} \quad \boldsymbol{E_1, E_2} \in \mathbb{R}^{\boldsymbol{N \times d}} \tag{9}$$

We can set a and B vectors as learnable parameters, so that we can realize data-driven graph structure learning.In the process of encoder calculation, we only add spatial mask, which can avoid introducing unnecessary information in the process of obtaining the representation of time sequence. It is worth noting that in the process of encoder calculation, we do not consider adding temporary mask, which can ensure a better representation of spatiotemporal sequence.

4.4. Spatiotemporal Decoder

In the calculation process of decoder, we do not use the common iterative decoder paradigm like transformer, because the iterative decoder will not only greatly reduce the efficiency of model training and inference, increase the time consumption, but also have the problem of error transmission in the iterative process. In order to solve the above problems at the same time, we use the generative decoder. Instead of choosing a specific flag as the token like what research in NLP usually set, we sample a *L*token long sequence in the input sequence, which is an earlier slice before the output sequence. It works as follows:

Firstly, we sample L long sequence in input sequence and concatenate it with target long sequence filled with 0:

$$\boldsymbol{\mathcal{X}_{decoder} = Concate(\mathcal{X}_L, 0)} \in \mathbb{R}^{\left(\boldsymbol{L_{sample} L_{target}}\right) \times \boldsymbol{N \times F}} \tag{10}$$

Then, we send $\mathcal{X}_{decoder}$ to complete the subsequent computation with spatial mask and temporal mask simultaneously. We use a upper triangular matrix to perform as a temporal mask, which can efficiently avoid leaving information from future points to historical points.

4.5. Others

After the calculation of the decoder, we send the hidden layer representation of the obtained spatiotemporal sequence into the full connection layer, that is, we get the spatiotemporal sequence value to be predicted. MSE loss is selected as the loss function:

$$Loss = \frac{1}{N+K} \sum_{n=1}^{N} \sum_{t=1}^{K} \left(\hat{x}_{n,t} - x_{n,t}\right)^2 \tag{11}$$

5. Experiments

5.1. Datasets

We use two crowd flow prediction data sets – Taxi-NYC and Bike-NYC. Taxi-NYC is obtained from NYC-TLC, and Bike-NYC is obtained from Citi-Bike[1]. They contain 60 days of trip records, in which the locations and times of the start and the end of a trip is included. We use the first 40 days

as training data and the rest 20 days as test data. Since these data sets are frequently adopted in crowd flow prediction researches, we adopt the general settings, including gird size, time interval, and thresholds.

Table 1. Datasets

Data sets	Taxi-NYC	Bike-NYC
Map size	16×12	14×8
Grid size	1km×1km	1km×1km
Time interval	30 mins	30mins
Features	inflow/outflow	inflow/outflow
Futures	12	12
Max	1409/318	262/274
Mean	114.0/146.1	33.1/32.6
Std	141.6/167.2	26.7/26.9
Time Range	1/1/2016-2/29/2016	8/1/2016-9/29/2016
Records	28.1 million	3.8 million

5.2. Baselines

We compare our model with following baseline models:

1.MLP: Multi-Layer Perceptron, three-layer fully connected network. 2. LSTM: Long-short term memory neural network. 3. STGCN[9]: Spatio-Temporal Graph Convolution Network, which utilizes graph convolution and 1D convolution to capture spatial dependencies and temporal correlations respectively.4.ASTGCN-r[2]:Attention based Spatial Temporal Graph Convolutional Networks, which utilize spatial and temporal attention mechanisms to model spatial-temporal dynamics respectively. In order to keep the fairness of comparison, only recent components of modeling periodicity are taken. 5. STGODE[12]: Spatial-Temporal Graph Ordinary Differential Equation Networks, which capture spatial-temporal dynamics through a tensor-based ordinary differential equation (ODE), as a result, deeper networks can be constructed and spatial-temporal features are utilized synchronously.

5.3. Metrics and Experimental Setting

Three kinds of evaluation metrics are adopted, including root mean squared errors(RMSE), mean absolute errors(MAE), and mean absolute percentage errors(MAPE).

All experiments are conducted on a Linux server (CPU: Intel(R) Core (TM) i9-11900K @ 3.50GHz, GPU: NVIDIA RTX 3090 24GB). The hidden dimensions of TCN blocks are set to 64, 32, 64, the learnable parameters' dimension which is decomposition of graph adjacent matrix are set to 10, the linear output layers are set to 2 layers.We train our model using Adam optimizer with a learning rate of 0.002. The batch size is 32 and the training epoch is 400.

5.4. Short-term Prediction

In this section, we evaluate the effectiveness of our model and other baselines in modeling complex and dynamic spatial temporal correlations by examining short-term prediction results. As shown in Table 2, in the two crowd flow prediction tasks, the performance of deep learning method is better than that of traditional neural network because of its complex structure.

Table 2. Short-term Prediction

	Bike-NYC			Taxi-NYC		
Models	RMSE	MAE	MAPE	RMSE	MAE	MAPE
MLP	10.67	8.21	26.49	24.10	14.90	21.88
LSTM	10.53	8.03	26.16	25.35	13.76	25.52
STGCN	9.57	6.81	25.87	21.39	13.31	**17.44**
ASTGCN	9.88	7.04	23.94	21.40	13.42	18.40
STGODE	9.68	6.95	22.01	21.30	14.20	17.46
STATTN	**9.38**	**6.65**	**21.72**	**21.05**	**13.14**	18.30

We can see from the results in the table above that the short-term prediction accuracy of the more complex in-depth learning model on the two data sets of travel demand prediction is higher than that of the simple MLP and LSTM models. The biggest difference between the following more complex deep learning models and MLP and LSTM models is that they often use graph neural network or graph attention mechanism to better capture the correlation between nodes in spatiotemporal sequence units.The key reason why ASTGCN performs better than STGCN is that ASTGCN integrates the dependence of temporal dimension and the dependence of spatial dimension for modeling. Compared with the structure in which STGCN models and calculates the two dimensions separately, it can implicitly capture the spatiotemporal correlation information. Compared with the above two deep learning models, STGODE has further improved the prediction accuracy. Because STGODE uses a tensor-based ordinary differential equation (ODE) to capture spatial-temporal simultaneously, the use of this structure can model the relationship between time and space dimensions at the same time. At the same time, the adoption of ODE greatly improves the calculation efficiency of the model and avoids the problem of over smoothing caused by the superposition of module layers. It is a model with excellent theoretical support and practical performance. The key reason why our model can achieve good results on two travel demand forecast data sets is that we use the spatiotemporal self-attention network which can capture the non-linear spatiotemporal dependence better than STGODE. Furthermore, we use data-driven or so-called dynamic graph to guide our model to aggregate information from spatial neighbours which has been proven more efficient in many previous research.

5.5. Long-term Prediction

We evaluate the effectiveness of our model and other baselines in modeling complex and dynamic spatial temporal correlations by examining long-term prediction results. We take 12 steps history spatiotemporal stamps as input to make 12 steps ahead prediction in this section. Since the performance of MLP and LSTM models in this task is not particularly good, the comparison and analysis with MLP and LSTM models will not be carried out in this chapter.

Table 3. Long-term Prediction

	Bike-NYC			Taxi-NYC		
Models	RMSE	MAE	MAPE	RMSE	MAE	MAPE
STGCN	13.89	9.01	32.03	40.46	22.63	25.83
ASTGCN	13.07	8.82	24.46	39.26	22.57	30.88
STGODE	12.89	8.41	**22.43**	34.99	19.74	28.30
STATTN	**12.45**	**8.22**	22.63	**31.85**	**18.02**	**20.65**

As shown in table 3, we observe the increased errors for all evaluated approaches, compared to their results of merely performing short-term predictions. Specifically, in New York taxi data, the RMSE of DSAN increased by 10.8, while the RMSE of STGCN, ASTGCN and STGODE increased by 19.07, 17.86 and 13.69 respectively. It shows that connecting each output directly to all inputs

through the design of STATTN decoder module is crucial to mitigating error propagation, compared to relying entirely on previously predicted outputs. By comparing the performance of the baseline model and our model on two data sets and three evaluation indicators, we can see that our model is basically in the leading position in the task of long-term prediction. Therefore, it can be explained that the STATTN model using the X mechanism can achieve a good effect and play a certain advantage in the work of long series modeling and long series output prediction. In comparison, STATTN can still perform a reliable spatial-temporal prediction without considering any assisting information.

5.6. Ablation Study

To demonstrate the effects of different components in STATTN, we evaluate the following variants on long-term prediction task:1.-T-Mask: STATTN without temporal mask. 2.-S-Mask: STATTN without spatial mask. 3.-TCN: STATTN without dilated temporal convolutional network to get spatial points' representation, which use the raw features of spatial points. 4.-Decoder: STATTN without decoder to generate the prediction result. Relatively, we use a fully-connected network is to transform the encoder's output to the final output, which is part of complete STATTN.

Table 4. Ablation Result

	Bike-NYC			Taxi-NYC		
Models	RMSE	MAE	MAPE	RMSE	MAE	MAPE
-T-Mask	14.52	9.12	26.34	41.32	22.78	26.97
-S-Mask	13.60	8.64	23.95	34.88	19.47	21.97
-TCN	14.38	8.75	25.61	39.97	21.70	23.70
-Decoder	12.76	8.51	23.65	32.18	19.01	21.30
STATTN	**12.45**	**8.22**	**22.63**	**31.85**	**18.02**	**20.65**

As shown in Table 4, the variants are more or less not as competent as STATTN. It is obvious from the results in the table that the RMSE of the model without Temporal-Mask on the Bike-NYC dataset is 14.52, which is obviously weaker than that of STGCN, and the same is true on the Taxi-NYC dataset. The main reason is that the spatiotemporal phenomenon contains a large number of inputs (there are more than 1000 taxi companies in New York), which exceeds the ability of attention calculation. At the same time, if the temporary mask is not added to the decoder structure, it is very likely that the predicted value at the current stage will be affected by the future stage. For other variants, the RMSE of STATTN without Spatial-Mask in Bike-NYC task is increased by 9.23%, because there is no indication of spatial location in the attention calculation process. In addition, the lack of TCN will make the model unable to obtain high-quality spatial point representation, which lead to 13.38% increased in RMSE in Bike-NYC task. However, the adoption of decoder can help the model to directly output the values of all time steps that need to be predicted and avoid the propagation of error. However, according to the experimental results, we can only see that the addition of decoder does not help us to improve the performance of the model to a great extent. On the bike NYC dataset, the RMSE error is only reduced by 2.4%. As a conclusion, we can see that the adoption of temporary mask and TCN has a great impact on the performance of our STATTN model, because these two structures are related to the indication of attention mechanism in the time dimension and the acquisition of high-quality spatial point representation.

6. Conclusions

In this work, we present a novel model for spatiotemporal traffic demand forecasting. Our model could capture spatiotemporal-dependencies simultaneously and effectively by a novel spatiotemporal self-attention network. Using Dilated TCN to generate spatial nodes' representation , using self-attention mechanism to capture spatiotemporal dependencies and generation-style decoder prevent the

error accumulation between inference. Detailed experiments and analysis reveal the strengths and weaknesses of previous models, which can demonstrate the excellent performance of STATTN.

7.Acknowledgments

This work was supported by National Key R&D Program of China (2019YFB2101804).

8.References

[1] Lin, H., Bai, R., Jia, W., Yang, X., & You, Y., 2020. Preserving dynamic attention for long-term spatial-temporal prediction. In: the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.San Diego, CA. 36-46.

[2] Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H., 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: the AAAI conference on artificial intelligence. Hilton Hawaiian Village, Honolulu, Hawaii, USA. 922-929.

[3] Wu, Y., & Tan, H., 2016. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. arXiv preprint arXiv:1612.01022.

[4] Wu, Yuankai, and Huachun Tan. "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework." arXiv preprint arXiv:1612.01022 (2016).

[5] Yuan, Z., Zhou, X., & Yang, T., 2018. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In: the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK. 984-992.

[6] Kim, S., Hong, S., Joh, M., & Song, S. K., 2017. Deeprain: Convlstm network for precipitation prediction using multichannel radar data. arXiv preprint arXiv:1711.02316.

[7] Wang, D., Yang, Y., & Ning, S., 2018. DeepSTCL: A deep spatio-temporal ConvLSTM for travel demand prediction. In:2018 international joint conference on neural networks. Rio de Janeiro,Brazil.1-8.

[8] Lin, Z., Li, M., Zheng, Z., Cheng, Y., & Yuan, C., 2020. Self-attention convlstm for spatiotemporal prediction. In: the AAAI Conference on Artificial Intelligence. New York, USA.11531-11538.

[9] Yu, B., Yin, H., & Zhu, Z., 2017. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875.

[10] Song, C., Lin, Y., Guo, S., & Wan, H., 2020. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In: the AAAI Conference on Artificial Intelligence. New York, USA.914-921.

[11] Li, M., & Zhu, Z., 2021. Spatial-temporal fusion graph neural networks for traffic flow forecasting. In: the AAAI conference on artificial intelligence. Vancouver, Canada.4189-4196.

[12] Fang, Z., Long, Q., Song, G., & Xie, K., 2021. Spatial-temporal graph ode networks for traffic flow forecasting. In: the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. Singapore. 364-373.

[13] Chen, R. T., Rubanova, Y., Bettencourt, J., & Duvenaud, D. K. (2018). Neural ordinary differential equations. Advances in neural information processing systems, 31.

[14] Xhonneux, L. P., Qu, M., & Tang, J., 2020. Continuous graph neural networks. In: International Conference on Machine Learning. New York, USA. 10432-10441.

[15] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y., 2017. Graph attention networks. arXiv preprint arXiv:1710.10903.

[16] Fang, X., Huang, J., Wang, F., Zeng, L., Liang, H., & Wang, H., 2020. Constgat: Contextual spatial-temporal graph attention network for travel time estimation at baidu maps. In: the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. San Diego, CA. 2697-2705.

[17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I., 2017. Attention is all you need. In: Advances in neural information processing systems. California, USA. 30.

[18] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W., 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In: the AAAI Conference on Artificial Intelligence. Vancouver, Canada.11106-11115.
[19] Song, C., Lin, Y., Guo, S., & Wan, H.,2020. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In: the AAAI Conference on Artificial Intelligence. New York, USA. 914-921.