Spatio-temporal multi-attention graph network for traffic forecasting

Qinzheng Li[§], Wenxing Zhu*
School of Control Science and Engineering, Shandong University, Jinan, China

ABSTRACT

Traffic forecasting is one of the most important problems in the areas of intelligent transportation system, and it is the key link. It plays a major role in transportation service and navigation. However, urban traffic has its own characteristics, and the complex traffic system is highly nonlinear and stochastic, which makes traffic forecasting a very difficult problem. Although many previous methods can make the high performance for predicting in traffic forecasting, the existing research has not fully utilized the influence of spatial and temporal characteristics on prediction. In this article, we put forward a new model called Spatio-Temporal Multi-Attention Graph Network. Taking into account the similar features of traffic flow every day and the interaction between road network structures, the model takes advantages of the internal dependence between the dynamic spatial network and the time dimension information to improve accuracy of forecasting. Experimental results show that our model is nicer over the others, which has good performance and gain more precision prediction accuracy.

Keywords: Traffic forecasting, deep learning, attention model, graph neural networks, spatiotemporal relationship

1. INTRODUCTION

With the increasing complexity of actual traffic problems, the theories and methods of traffic prediction are still constantly renewal and development. Traffic forecasting is a very impotant issue in traffic control, it refers to the analysis of a large amount of historical data to predict the future traffic conditions as much as possible to help with traffic decisions to better control traffic and reduce traffic congestion.

Long-term traffic flow forecasting is a very challenging task, which is determined by its high complexity, nonlinear time correlation, dynamic spatial correlation, and long-term accumulation of errors. With the development of science and technology, we can now obtain a large number of traffic time series data from the information collection equipment on expressways, which provides a good foundation for traffic big data forecasting. In the field of time series forecasting, traditional time-sequence analysis such as Autoregressive Integrated Moving Average (ARIMA)[1] is still very popular, but it is difficult to deal with unstable and non-linear data. Recently years, the rapid growth of deep learning model has brought more possibilities, a lot of researchers have begun to use convolutional neural networks (CNN) in the feature extraction, but losing sight of spatiotemporal correlation. Defferrard et al.[2] looks for potential data to find relations by Graph Convolutional Networks (GCN), but only for undirected graphs. Li et al.[3] skilfully applies diffusion convolution to extract spatial features well, but the extraction of temporal features is not perfect.

To attack the above problems, we propose a Spatio-Temporal Multi-Attention Graph Network (**STMAGN**), which has an appropriate architecture and gets good results. We extract the features of historical traffic data through the encoder, and the decoder uses the output sequences of previous structure. In order to reduce the impact of error propagation, we add a conversion layer before decoding. In this work, we use two mechanisms of attention to model the connection between time and space and gating them together to fuse information features. The multi-head attention is to discover the inherent correlation relationship of the time series from different angles. The model effectively captures the dynamic features and improves the prediction accuracy.

2. PRELIMINARY

We define the road network structure as a directed graph $\mathcal{G} = (\mathcal{V}, \varepsilon, W)$. Here, $\mathcal{V}$ represents a collection of all nodes $|\mathcal{V}| = N$, indicating the connectivity among nodes. W is the adjacency matrix representing the relationship among roads.

The target of traffic forecasting is to use a large amount of historical data to predict various traffic parameters in the future, which is a standing dish question. Assuming we now have the information collected by the sensors on the road,

§202034861@mail.sdu.edu.cn; *zhuwenxing@sdu.edu.cn

International Conference on Computer Application and Information Security (ICCAIS 2021),
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12260, 122600F · © The Authors.
Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2637523

we use $X_t \in R^{N\times C}$ to represents the observed traffic flow information, where C represents various status information of the road.

Given the observations of historical P time steps $\mathrm{X} = \left(X_{t_1}, X_{t_2}, \cdots, X_{t_p}\right)$ at N vertices, our goal is to learn a sophisticated function $F(\cdot)$ to connect the future Q time steps with the historical P time steps:

$$[X_{t_{p+1}}, X_{t_{p+2}}, \cdots, X_{t_{p+1}}] = F\left([X_{t_1}, X_{t_2}, \cdots, X_{t_p}]\right) \tag{1}$$

3. SPATIO-TEMPORAL MULTI-ATTENTION GRAPH NETWORK

Figure 1 presents the whole structure of the STMAGN model mentioned in this article.

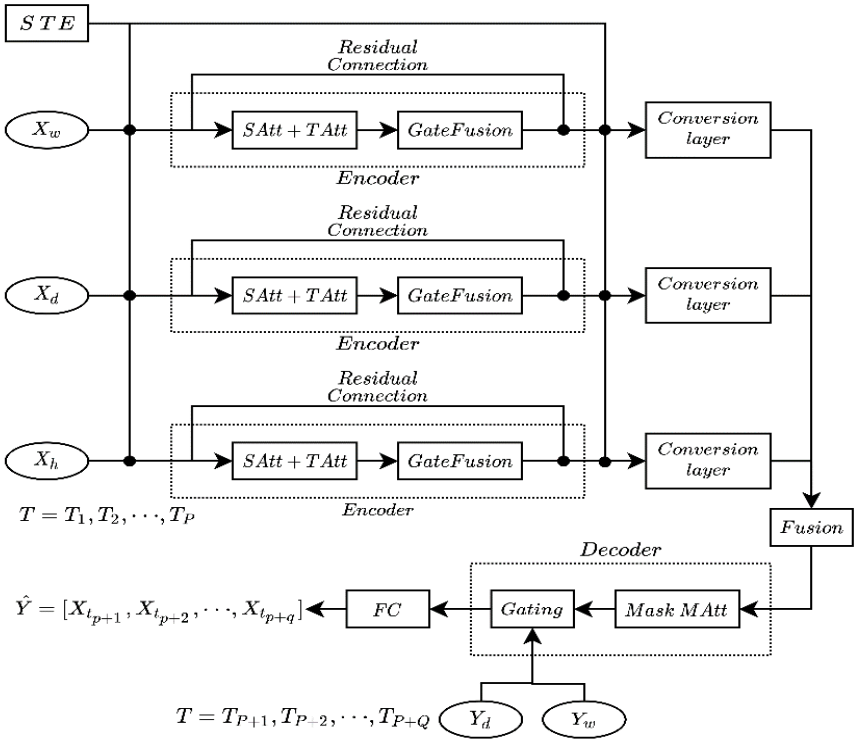

Figure 1. Spatio-temporal multi-attention graph network.

It is composed of an encoder-decoder structure and conversion layer. The encoder includes a spatio-temporal attention module with residual connection[4] and an information fusion structure. The decoder includes a mask multi-head attention mechanism and gating structure. The conversion layer between them is responsible for converting the features extracted by the encoder into the decoder.

Suppose we want to predict the data for a specific period of time, we will extract the time data of week, day and hour at the same time for modelling respectively to fully capture the periodicity of traffic flow. These inputs will be encoded by the encoder and transmitted to the conversion layer, and finally decoded by the decoder to obtain the output.

3.1 Spatial-temporal embedding

In practice, the evolution of traffic state will be affected by the basic traffic network structure, so it is necessary to build the network structure and input it into the prediction model. As shown in Figure 2, we model spatial dependence by associating traffic flow with diffusion process, which clearly captures the randomness of road network.

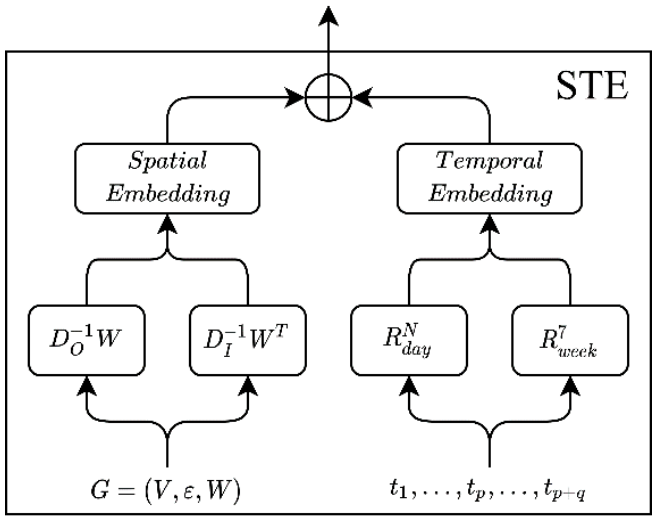

Figure 2. Spatio-temporal embedding.

The characteristic of the diffusion process is a random walk on graph, and the restart probability is α, with a probability matrix of information transfer $D_O^{-1}W^3$. $D_O = diag(W)$. The matrix will restrain itself to an equilibrium probability condition $P \in R^{N\times N}$. Teng et al.[5] pointed out that the ultimate stable state can be obtained by the following:

$$P = \sum_{k=0}^{\infty} \alpha\ (1-\alpha)^k (D_O^{-1}W)^k \tag{2}$$

where k is the diffusion steps. In this work, we take the truncation of finite k steps and model the spatial dependence by bidirectional diffusion. Hence, we can express spatial embedding in the following form:

$$e_{v_i}^S = \sum_{k=0}^{\infty} \theta_k ((D_O^{-1}W)^k + (D_I^{-1}W^T)^k) \tag{3}$$

The spatial embedding only provides static representation and cannot represent the dynamic correlation. Therefore, we put forward another way, which encodes time dimension as a vector. We divide a day into N parts, encoded the time steps into R^N and R^7 by one-hot coding and spliced into R^{N+7}, represented as $e_{t_j}^T$[6].

In our model, we both unify these features into R^D through a fully connected neural module and fuse them as spatio-temporal embedding (STE). Therefore, the STE includes both road network structure and time features.

3.2 Multi-head attention

Since the attention mechanism was proposed, it has been applied extensively in many fields. It can find out the relationship between them according to the raw data and extract the most important features. Multi-head Attention is to calculate the attention of the data in different subspace with the total number of parameters unchanged, and the last step is to merge the attention information in different subspace[7]. The dimension of each vector is reduced by this way when calculating the attention of each head and the over-fitting phenomenon is also avoided; because attention has different parameters in different subspace, Multi-head Attention looks for the correlation between sequences relations from different angles in fact.

For the next state of node i at time t, we update it with the sum of the corresponding weights of all nodes can be expressed as follow:

$$H^l = \sum \alpha \cdot H^{l-1} \tag{4}$$

α is the attention score indicating the significance of node,H^{l-1} indicates the last hidden state and H^l indicates the current state. Adopting the scaled dot-product approach[7] to learn attention score.

$$\frac{<f_1(H^{l-1}\|(e_{v_i}+e_{t_j}))>,f_2(H^{l-1}\|(e_{v_i}+e_{t_j}))>}{\sqrt{d}} \tag{5}$$

where || indicates the splicing process, $\left(e_{v_i} + e_{t_j}\right)$ represents spatio-temporal embedding vector, f_1 and f_2 are an activation function with two different parameters, d is the dimension after vector splicing. Then, λ^k is normalized to α by SoftMax[8], by splicing K attention mechanisms with different learning styles, we can get:

$$H^l = \|_{k=1}^{K} \sum \alpha \cdot f_3(H^{l-1}) \tag{6}$$

f_3 is another activation function. Therefore, we successfully capture the inner spatial relationship among nodes.

3.3 Gate fusion

In order to further integrate the spatio-temporal relationship, we designed a gated fusion module to adaptively fuses spatiotemporal information as shown in Figure 3.

After the input passes through the spatial and temporal attention mechanism, the output is represented as H_S^l and H_T^l, H_S^l and H_T^l merge into:

$$r^l = \sigma(W_1(H_S^l \parallel H_T^l) + b_1)$$
$$H^l = (1 - r^l)H_S^l + r^l H_T^l \tag{7}$$

where W_1 and b_1 are different learnable parameters, σ is the sigmoid function, H_S^l and H_T^l is the result of spatio and temporal attention.

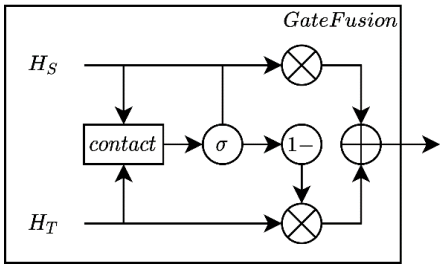

Figure 3. Fuse spatio attention and temporal attention information together.

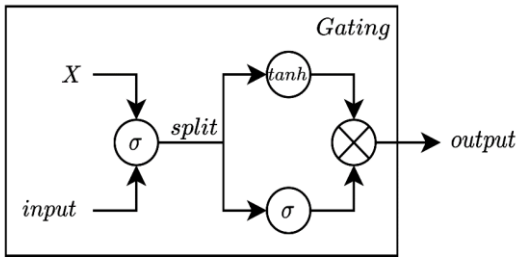

Figure 4. Control information transmission.

3.4 Gating

In order to make the final results obtain the characteristics of long-time period, we introduce a gating mechanism to correct the attention results as shown in Figure 4, so as to reduce the prediction error.

In this way, the network model can not only remember the information of the past, but also selectively forget some inessential information and shape long-term relationships, the calculation process is as follows:

$$z_1, z_2 = split\left(\sigma(W_2(X \parallel input) + b_2)\right)$$
$$output = tanh(z_1) \otimes \sigma(z_2) \tag{8}$$

where split means separating the output results, $tanh$ is another activation function, the gating mechanism can combine the obtained output with historical information to get more accurately results.

4. EXPERIMENTS

4.1 Datasets

We used our model to make traffic predictions on real data set PeMS-bay. In this data set, we take the traffic speed every five minutes and normalize the data to the interval $[0, 1]$. In order to build the road network, we calculate the paired road distance and use the threshold to construct adjacency matrix[9] $W_{ij} = exp\left(-dist\left(v_i, v_j\right)^2 \sigma^{(-2)}\right)$ if $\text{dist}\left(v_i, v_j\right) \leq \delta$, otherwise 0, where W_{ij} represents the weight of the adjacency matrix. $dist\left(v_i, v_j\right)$ indicates the distance between sensors. σ is the standard deviation indicates the degree of dispersion of the distance, and δ is the limit value remove unnecessary data.

4.2 Result analysis

In order to make the model results easier to understand, we visualize the experiments results. Figure 5 shows experimental results on actual traffic data set of the model. From these figures, we conclude that when the average road speed fluctuates little, the model can generate a relatively smooth curve to fit the road traffic conditions. Even if the road conditions change suddenly, the model can capture this change according to the spatio-temporal characteristics and generate more accurate prediction curve.

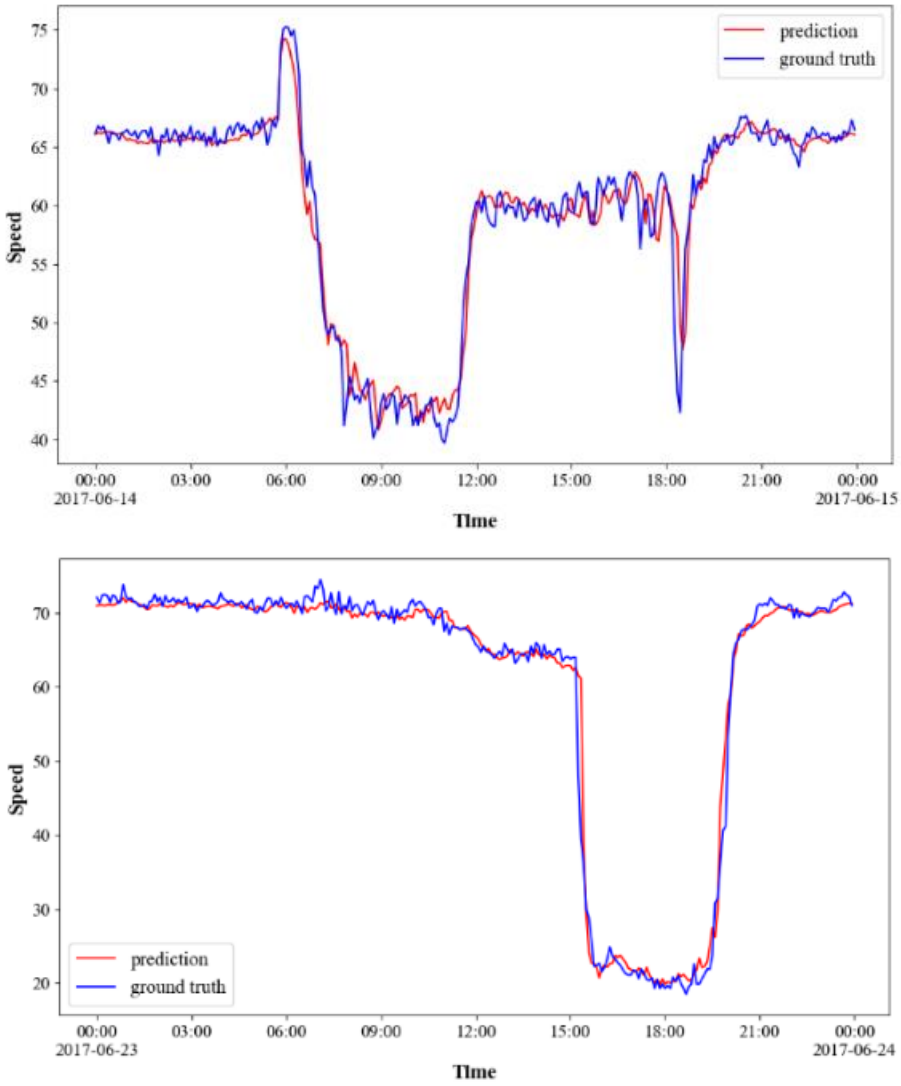

Figure 5. Result on real data of PeMS-bay.

4.3 Baselines and comparison

We compare our model with other baselines. Table 1 shows the average results of traffic forecasting estimated performance in the next one hour. It can be seen that our STMAGN gets a good form in all aspects of all evaluation indicators, especially in the long-term prediction stage, it has achieved far better results than other models. In addition, we can observe that the results of traditional sequences approaches are not perfect in general, which shows that these methods have limited capacity for increasingly complex transportation systems. Through comparison, the methods on account of deep learning have generally achieved good results.

Table 1. Comparison of results between STMAGN and the others.

	Horizon 3			Horizon 6			Horizon 12		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
PEMS-BAY									
ARIMA	1.61	3.30	3.50%	2.32	4.78	5.42%	3.40	6.51	8.32%
SVR	1.87	3.60	3.81%	2.50	5.17	5.52%	3.29	7.09	8.02%
VAR	1.72	3.16	3.60%	2.30	4.26	5.02%	2.94	5.46	6.50%
FNN	2.22	4.40	5.20%	2.28	4.65	5.41%	2.46	4.97	5.91%
FC-LSTM	2.07	4.21	4.80%	2.21	4.55	5.22%	2.35	4.95	5.72%
STGCN	**1.37**	2.96	**2.89%**	1.81	4.25	4.17%	2.49	5.70	5.81%
DCRNN	**1.37**	**2.95**	2.91%	1.75	3.99	3.91%	2.07	4.72	4.92%
ASTGCN	1.52	3.13	3.22%	2.01	4.27	4.48%	2.61	5.42	6.00%
STMAGN	1.48	3.02	3.42%	**1.73**	**3.73**	**4.14%**	**1.93**	**4.19**	**4.79%**

4.4 Effectiveness of each module

In order to study the impact of each module, we evaluate in three different ways, like removing conversion layer or STE or gating from the model. Figure 6 shows the impact after removing these components, thus proving the effectiveness of each module.

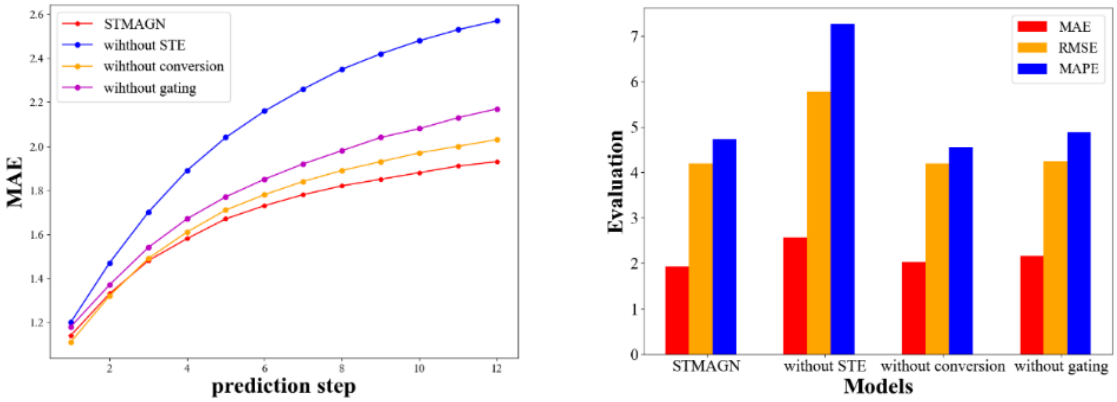

Figure 6. Effectiveness of each part of the model.

The experimental results show that removing any module has an established impact on the final prediction accuracy, especially the accuracy of modules without STE is dropping fast.

5. CONCLUSION

In our work, we put forward a spatio-temporal multi-attention graph network (STMAGN) to predict traffic situation. Specifically, we use a spatial and temporal attention mechanism to simulate intricate traffic situations, and propose bidirectional diffusion convolution and one-hot encoding to capture the dynamic spatio-temporal features more effectively and integrate them together. In addition to the above methods, we also apply the conversion layer to avoid error accumulation and the gating mechanism to obtain more accurate output. Experiments on real data set show that the prediction accuracy of model has good prediction accuracy.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (Grant No.61773243), Major Technology Innovation Projects of Shandong Province (Grant No. 2019TSLH0203) and the National Key Research and Development Program of China (Grant No. 2020YFB1600501).

REFERENCES

[1] Williams, B. M. and Hoel, L. A., "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," Journal of Transportation Engineering, 129(6), 664-672 (2003).
[2] Defferrard, M., Bresson, X. and Vandergheynst, P., "Convolutional neural networks on graphs with fast localized spectral filtering," NIPS, 3837-3845 (2016).
[3] Li, Y., Yu, R., Shahabi, C. and Liu, Y., "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," Proc. of ICLR, (2017).
[4] He, K., Zhang, X., Ren, S. and Sun, J., "Deep residual learning for image recognition," 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 770-778 (2016).
[5] Teng, S. H., "Scalable algorithms for data and network analysis," Foundations and Trends® in Theoretical Computer Science, 12(1-2), 1-274 (2016).
[6] Zheng, C., Fan, X., Wang, C. and Qi, J., "GMAN: A graph multi-attention network for traffic prediction," Proc. of the AAAI Conf. on Artificial Intelligence, 34(1), 1234-1241 (2020).
[7] Vaswani, A., Shazeer, N., Parmar, N., et al., "Attention is all you need," NeurIPS, 5998-6008 (2017).
[8] Wang, X., Ma, Y., Wang, Y, Jin, W. and Yu, J., "Traffic flow prediction via spatial temporal graph neural network," Web Conf., 1082-1092 (2020).
[9] Shuman, D. I., Narang, S. K., Frossard, P., et al., "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," IEEE Signal Processing Magazine, 30(3), 83-98 (2013).