Learning Dynamics and Heterogeneity of Spatial-Temporal Graph Data for Traffic Forecasting

ASTGNN+Transformer

Paper by Shengnan Guo Youfang Lin Huaiyu Wan Xiucheng Li Gao Cong
Presentation by Tobramycin

NCUT

2022.6

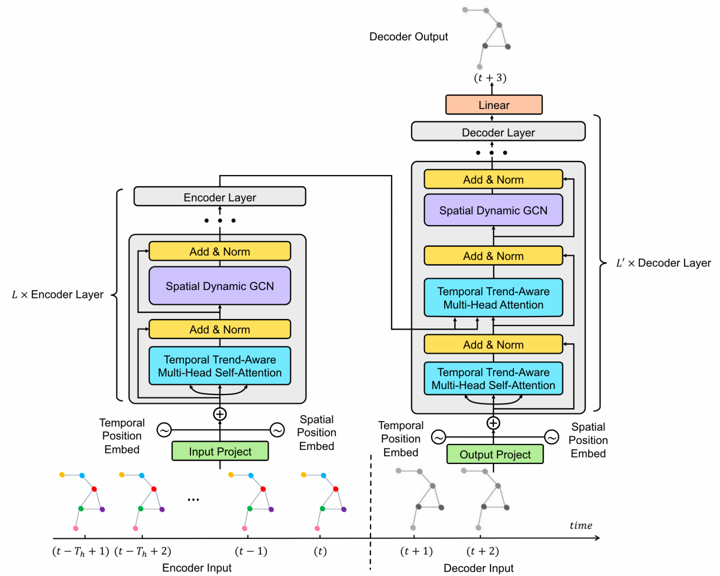
Contents

Introduction

Introduction

ASTGNN

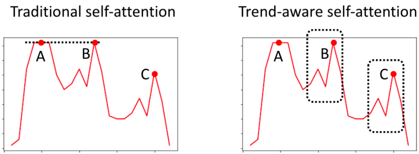
Framework

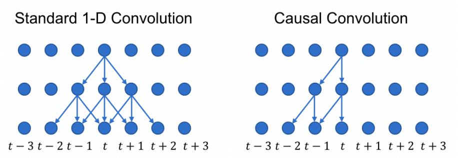


Decoder Output
$(t+3)$
Linear
Decoder Layer
Add & Norm
Spatial Dynamic GCN
Add & Norm
Temporal Trend-Aware Multi-Head Attention
Add & Norm
Temporal Trend-Aware Multi-Head Self-Attention
$L' \times$ Decoder Layer
Encoder Layer
Add & Norm
Spatial Dynamic GCN
Add & Norm
Temporal Trend-Aware Multi-Head Self-Attention
$L \times$ Encoder Layer
Temporal Position Embed
Spatial Position Embed
Input Project
Temporal Position Embed
Output Project
Spatial Position Embed
time
$(t-T_h+1)$ $(t-T_h+2)$ ... $(t-1)$ (t)
Encoder Input
$(t+1)$ $(t+2)$
Decoder Input

Temporal trend-aware multihead attention

$$
\begin{aligned}
\operatorname{MHSelfAttention}(\mathrm{Q}, \mathrm{K}, \mathrm{V}) &= \oplus\left(\operatorname{head}_1, \ldots, \text{ head }_h\right) W^O \\
\operatorname{head}_j &= \operatorname{Attention}\left(\mathrm{Q} W_j^Q, \mathrm{K} W_j^K, \mathrm{V} W_j^V\right) \\
\operatorname{TrSelfAttention}(\mathrm{Q}, \mathrm{K}, \mathrm{V}) &= \oplus\left(\operatorname{Trhead}_1, \ldots, \operatorname{Trhead}_h\right) W^O \\
\operatorname{Trhead}_j &= \operatorname{Attention}\left(\Phi_j^Q \star \mathrm{Q}, \Phi_j^K \star \mathrm{K}, \mathrm{V} W_j^V\right)
\end{aligned}
$$





Positional Embedding

Temporal Position Embedding

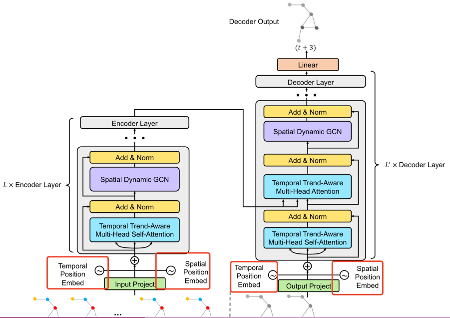
$$E_{\mathrm{TP}}(t, 2d) = \sin\left(t / 10000^{2d/d_{\text{model}}}\right) \tag{1}$$

$$E_{\mathrm{TP}}(t, 2d+1) = \cos\left(t / 10000^{2d/d_{\text{model}}}\right) \tag{2}$$

Spatial Position Embedding use Laplacian Smoothing

$$E_{\mathrm{SP}} = \tag{3}$$

0	1	2	...	510	511
sin(f(0, 0))	cos(f(0, 1))	sin(f(0, 2))	...	sin(f(0, 510))	cos(f(0, 511))
sin(f(1, 0))	cos(f(1, 1))	sin(f(1, 2))	...	sin(f(1, 510))	cos(f(1, 511))
sin(f(2, 0))	cos(f(2, 1))	sin(f(2, 2))	...	sin(f(2, 510))	cos(f(2, 511))
sin(f(3, 0))	cos(f(3, 1))	sin(f(3, 2))	...	sin(f(3, 510))	cos(f(3, 511))
sin(f(4, 0))	cos(f(4, 1))	sin(f(4, 2))	...	sin(f(4, 510))	cos(f(4, 511))
sin(f(5, 0))	cos(f(5, 1))	sin(f(5, 2))	...	sin(f(5, 510))	cos(f(5, 511))
sin(f(6, 0))	cos(f(6, 1))	sin(f(6, 2))	...	sin(f(6, 510))	cos(f(6, 511))
sin(f(7, 0))	cos(f(7, 1))	sin(f(7, 2))	...	sin(f(7, 510))	cos(f(7, 511))
sin(f(8, 0))	cos(f(8, 1))	sin(f(8, 2))	...	sin(f(8, 51[illegible]	[illegible]
sin(f(9, 0))	cos(f(9, 1))	sin(f(9, 2))	...	sin(f(9, 510))	cos(f(9, 511))



Spatial Dynamic GCN

Normal GCN

$$\mathrm{GCN}\left(\mathrm{Z}_t^{(l-1)}\right)=\sigma\left(\mathrm{A}\mathrm{Z}_t^{(l-1)}\mathrm{W}^{(l)}\right) \tag{4}$$

$$\mathrm{A}=\begin{cases}\widetilde{\mathrm{D}}^{-\frac{1}{2}}\widetilde{\mathrm{A}}\widetilde{\mathrm{D}} & \text{undirected graph}\\ \widetilde{\mathrm{D}}^{-1}\widetilde{\mathrm{A}}, & \text{directed graph}\end{cases} \tag{5}$$

Dynamic GCN

$$\mathrm{X}_t^{(l)}=\mathrm{DGCN}\left(\mathrm{Z}_t^{(l-1)}\right)=\sigma\left(\left(\mathrm{A}\odot\mathrm{S}_t\right)\mathrm{Z}_t^{(l-1)}\mathrm{W}^{(l)}\right) \tag{6}$$

Here, spatial correlation weight matrix

$$\mathrm{S}_t=\operatorname{softmax}\left(\frac{\mathrm{Z}_t^{(l-1)}{\mathrm{Z}_t^{(l-1)}}^T}{\sqrt{d_{\text{model}}}}\right)\in\mathbb{R}^{N\times N} \tag{7}$$

EXPERIMENTS

Datasets

TABLE 1
Dataset Description.

Data type	Datasets	# of sensors	Time range
Highway traffic flow	PEMS03	358	09/01/2018 - 11/30/2018
	PEMS04	307	01/01/2018 - 02/28/2018
	PEMS07	883	05/01/2017 - 08/31/2017
	PEMS08	170	07/01/2016 - 08/31/2016
Metro crowd flow	HZME	80	01/01/2019 - 01/26/2019

Baselines

- VAR: Vector Auto-Regression is an advanced time series model, which captures the pairwise relationships among multiple time series.
- SVR : Support Vector Regression utilizes a linear support vector machine to perform regression.
- LSTM [60]: Long Short-Term Memory network, a special RNN model.
- DCRNN [6]: Diffusion Convolutional Recurrent Neural Network employs diffusion graph convolutional networks and GRU based on seq2seq to predict traffic graph series data.
- STGCN [45]: Spatial-Temporal Graph Convolutional Network uses ChebNet in the spatial dimension and 2D convolutional networks in the temporal dimension to model the correlations in spatial-temporal graph data.
- ASTGCN [11]: Attention Based Spatial-Temporal Graph Convolutional Networks designs spatial attention and temporal attention mechanisms to model spatial and temporal dynamics.
- Graph WaveNet [10]: Graph WaveNet combines graph convolutions with temporal convolutions to capture spatial-temporal dependencies.
- STSGCN [12]: Spatial-Temporal Synchronous Graph Convolutional Network proposes a new kind of convolution operation to capture

Result

TABLE 3
Performance comparison on the four highway traffic flow datasets.
(the best results are in bold and * denotes the second-best results. † denotes our models.)

Baseline methods		VAR	SVR	LSTM	DCRNN	STGCN	ASTGCN(r)	Graph WaveNet	STSGCN	**ASTGNN**†	**ASTGNN(p)**†
Datasets	Metrics										
PEMS03	MAE	21.08	22.01 ± 0.07	20.62 ± 0.19	18.39 ± 0.17	18.28 ± 0.39	17.85 ± 0.45	14.79 ± 0.08	17.51 ± 0.13	14.78* ± 0.05	**14.55 ± 0.07**
	RMSE	34.75	35.28 ± 0.08	33.54 ± 0.34	30.56 ± 0.17	30.73 ± 0.78	29.88 ± 0.65	25.51 ± 0.17	29.05 ± 0.40	25.00* ± 0.18	**24.96 ± 0.31**
	MAPE (%)	22.35	22.93 ± 1.09	28.94 ± 2.76	20.22 ± 2.83	17.52 ± 0.32	17.65 ± 0.79	14.32* ± 0.24	16.92 ± 0.22	14.79 ± 0.22	**13.66 ± 0.14**
PEMS04	MAE	23.75	28.66 ± 0.01	26.81 ± 0.31	23.65 ± 0.04	22.27 ± 0.18	22.42 ± 0.19	19.36 ± 0.02	21.08 ± 0.14	18.60* ± 0.06	**18.44 ± 0.08**
	RMSE	36.66	44.59 ± 0.02	40.74 ± 0.17	37.12 ± 0.07	35.02 ± 0.19	34.75 ± 0.19	31.72 ± 0.13	33.83 ± 0.27	**30.91 ± 0.22**	31.02* ± 0.18
	MAPE (%)	18.09	19.15 ± 0.04	22.33 ± 1.60	16.05 ± 0.10	14.36 ± 0.12	15.87 ± 0.36	13.31 ± 0.19	13.88 ± 0.07	**12.36 ± 0.11**	12.37* ± 0.08
PEMS07	MAE	101.20	32.97 ± 0.98	29.71 ± 0.09	23.60 ± 0.05	27.41 ± 0.45	25.98 ± 0.78	21.22 ± 0.24	23.99 ± 0.14	20.62* ± 0.12	**19.26 ± 0.17**
	RMSE	155.14	50.15 ± 0.15	45.32 ± 0.27	36.51 ± 0.05	41.02 ± 0.58	39.65 ± 0.89	34.12 ± 0.18	39.32 ± 0.31	34.00* ± 0.21	**32.75 ± 0.25**
	MAPE (%)	39.69	15.43 ± 1.22	14.14 ± 1.00	10.28 ± 0.02	12.23 ± 0.38	11.84 ± 0.69	9.07 ± 0.20	10.10 ± 0.08	8.86* ± 0.10	**8.54 ± 0.19**
PEMS08	MAE	22.32	23.25 ± 0.01	22.19 ± 0.13	18.22 ± 0.06	18.04 ± 0.19	18.86 ± 0.41	15.07 ± 0.17	17.10 ± 0.04	15.00* ± 0.35	**12.72 ± 0.09**
	RMSE	33.83	36.15 ± 0.02	33.59 ± 0.05	28.29 ± 0.09	27.94 ± 0.18	28.55 ± 0.49	23.85* ± 0.18	26.83 ± 0.06	24.70 ± 0.53	**22.60 ± 0.13**
	MAPE (%)	14.47	14.71 ± 0.16	18.74 ± 2.79	11.56 ± 0.04	11.16 ± 0.10	12.50 ± 0.66	9.51 ± 0.22	10.90 ± 0.05	9.50* ± 0.11	**8.78 ± 0.20**

Result

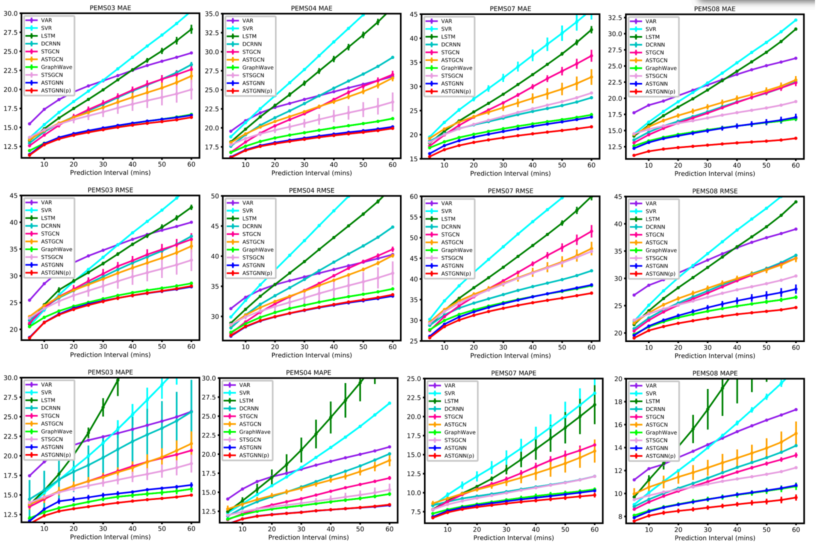


Fig. 6. Performance changes of different methods on the four datasets as the prediction interval increases.

Result

TABLE 5
Performance comparison on the metro crowd flow dataset.
(the best results are in bold and * denotes the second-best results. † denotes our models.)

Baseline methods	MAE		RMSE		MAPE (%)	
Metrics	inflow	outflow	inflow	outflow	inflow	outflow
VAR	17.65	22.35	28.10	37.96	58.07	96.68
SVR	21.94 ± 0.0173	25.59 ± 0.1187	40.73 ± 0.0205	50.07 ± 0.1732	49.40 ± 0.0679	91.71 ± 3.1788
LSTM	22.53 ± 0.5089	26.18 ± 0.3193	39.33 ± 0.3506	48.91 ± 0.4458	60.12 ± 2.4417	103.06 ± 8.5229
DCRNN	12.25 ± 0.1316	18.02 ± 0.1579	20.91 ± 0.3290	31.45 ± 0.3879	25.53 ± 0.3811	**66.98± 1.6475**
STGCN	12.88 ± 0.2810	19.12 ± 0.2312	22.86 ± 0.3884	33.12 ± 0.3596	29.66 ± 1.4970	73.66 ± 1.4909
ASTGCN	13.10 ± 0.4733	19.35 ± 0.5071	23.23 ± 0.8127	33.20 ± 1.0714	33.29 ± 3.6336	88.75 ± 3.9984
Graph WaveNet	11.20* ± 0.1116	17.50* ± 0.1234	19.73* ± 0.4610	**30.65 ± 0.4059**	23.75* ± 0.7082	73.65 ± 2.7181
STSGCN	12.85 ± 0.1006	18.74 ± 0.1289	23.20 ± 0.3773	33.12 ± 0.4260	28.02 ± 0.1947	76.85 ± 1.0094
ASTGNN†	11.46 ± 0.0841	17.94 ± 0.1093	20.84 ± 0.2498	31.91 ± 0.3163	24.42 ± 0.3029	72.46 ± 2.4177
ASTGNN(p)†	**10.94 ± 0.0393**	**17.47 ± 0.0310**	**18.89 ± 0.1081**	30.78* ± 0.0774	**23.33 ± 0.1400**	70.52*± 0.2739

Conclusion

Conclusion

Thank you!