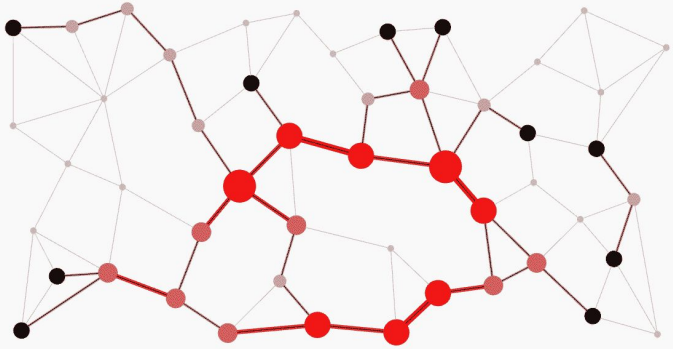
Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting

Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, Huaiyu Wan



[Image source]

The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)

Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting

Shengnan Guo,[1,2] **Youfang Lin,**[1,2,3] **Ning Feng,**[1,3] **Chao Song,**[1,2] **Huaiyu Wan**[1,2,3]*
[1]School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China
[2]Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing, China
[3]CAAC Key Laboratory of Intelligent Passenger Service of Civil Aviation, Beijing, China
{guoshn, yflin, fengning, chaosong, hywan}@bjtu.edu.cn*

Abstract

Forecasting the traffic flows is a critical issue for researchers and practitioners in the field of transportation. However, it is very challenging since the traffic flows usually show high nonlinearities and complex patterns. Most existing traffic flow prediction methods, lacking abilities of modeling the dynamic spatial-temporal correlations of traffic data, thus cannot yield satisfactory prediction results. In this paper, we propose a novel attention based spatial-temporal graph convolutional network (ASTGCN) model to solve traffic flow forecasting problem. ASTGCN mainly consists of three independent components to respectively model three temporal properties of traffic flows, i.e., recent, daily-periodic and weekly-periodic dependencies. More specifically, each component contains two major parts: 1) the spatial-temporal attention mechanism to effectively capture the dynamic spatial-temporal correlations in traffic data; 2) the spatial-temporal convolution which simultaneously employs graph convolutions to capture the spatial patterns and common standard convolutions to describe the temporal features. The output of the three components are weighted fused to generate the final prediction results. Experiments on two real-world datasets from the Caltrans Performance Measurement System (PeMS) demonstrate that the proposed ASTGCN model outperforms the state-of-the-art baselines.

Introduction

Recently, many countries are committed to vigorously develop the Intelligent Transportation System (ITS) (Zhang et al. 2011) to help for efficient traffic management. Traffic forecasting is an indispensable part of ITS, especially on the highway which has large traffic flows and fast driving speed. Since the highway is relatively closed, once a congestion occurs, it will seriously affect the traffic capacity. Traffic flow is a fundamental measurement reflecting the state of the highway. If it can be predicted accurately in advance, according to this, traffic management authorities will be able to guide vehicles more reasonably to enhance the running efficiency of the highway network.

Highway traffic flow forecasting is a typical problem of spatial-temporal data forecasting. Traffic data are recorded at fixed points in time and at fixed locations distributed

*Corresponding author: hywan@bjtu.edu.cn

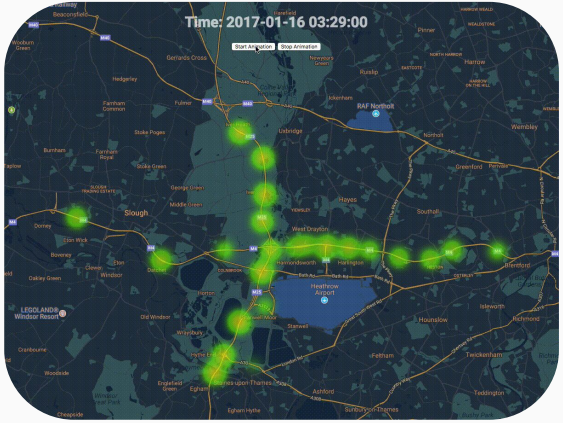
Figure 1: The spatial-temporal correlation diagram of traffic flow

in continuous space. Apparently, the observations made at neighboring locations and time stamps are not independent but dynamically correlated with each other. Therefore, the key to solve such problems is to effectively extracting the spatial-temporal correlations of data. Fig. 1 demonstrates the spatial-temporal correlations of traffic flows (also can be vehicle speed, lane occupancy, etc.). The bold line between two points represents their mutual influence strength. The darker the color of line is, the greater the influence is. In the spatial dimension (Fig. 1(a)), we can find that different locations have different impacts on A and even a same location has varying influence on A as time goes by. In the temporal dimension (Fig. 1(b)), the historical observations of different locations have varying impacts on A's traffic states at different times in the future. In conclusion, the correlations in traffic data on the highway network show strong dynamics in both the spatial dimension and temporal dimension. How to explore nonlinear and complex spatial-temporal data to discover its inherent spatial-temporal patterns and to make accurate traffic flow predictions is a very challenging issue.

Fortunately, with the development of the transportation industry, many cameras, sensors and other information collection devices have been deployed on the highway. Each

Motivation

- Intelligent Transportation System (ITS)
- Forecasting traffic flows are essential part of ITS
- Existing methods lack ability to capture dynamic correlations in spatial-temporal (ST) traffic data
- Key to solve the forecasting challenge lies in extracting ST correlations effectively from the data

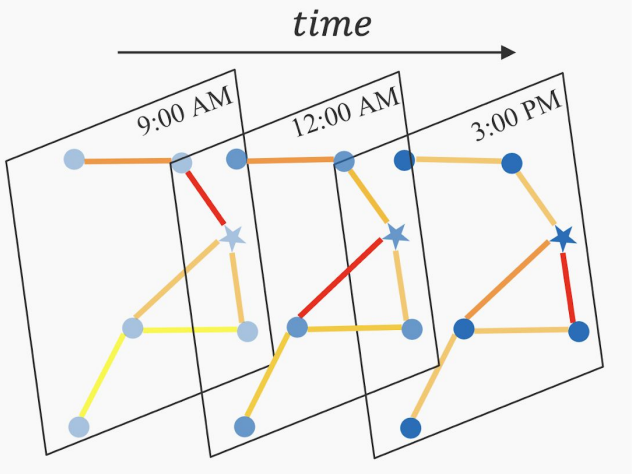


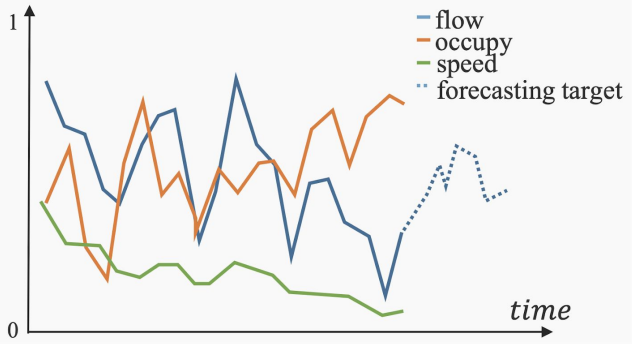
[Image source]

Preliminaries

Traffic Networks

- Traffic data is recorded
  - at fixed points in time
  - at fixed locations
- *G = V, E, A*
- Every sensor detects *F* features
  - at equal sampling frequencies





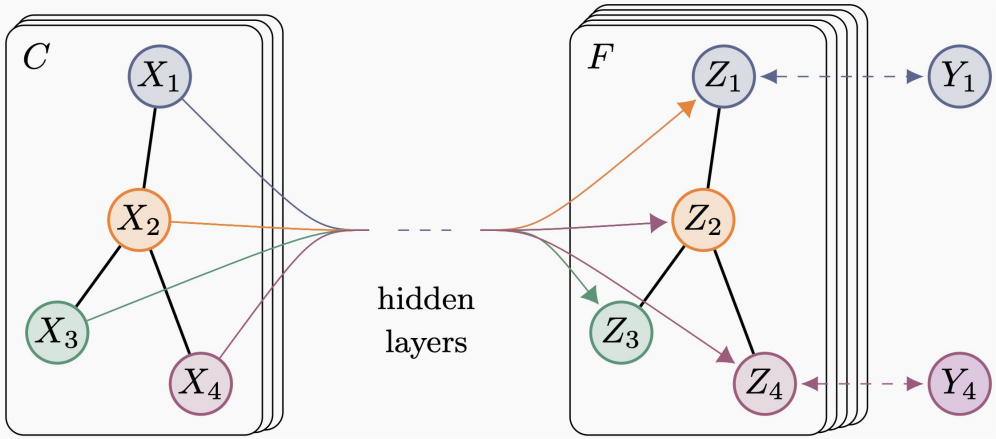
Related work

- **Statistical methods**
  - HA
  - ARIMA
  - VAR
- **Classical Machine Learning**
  - KNN
  - SVM
- **Deep Learning**
  - ST-ResNet
  - GeoMAN
  - DMVST-Net

NTNU

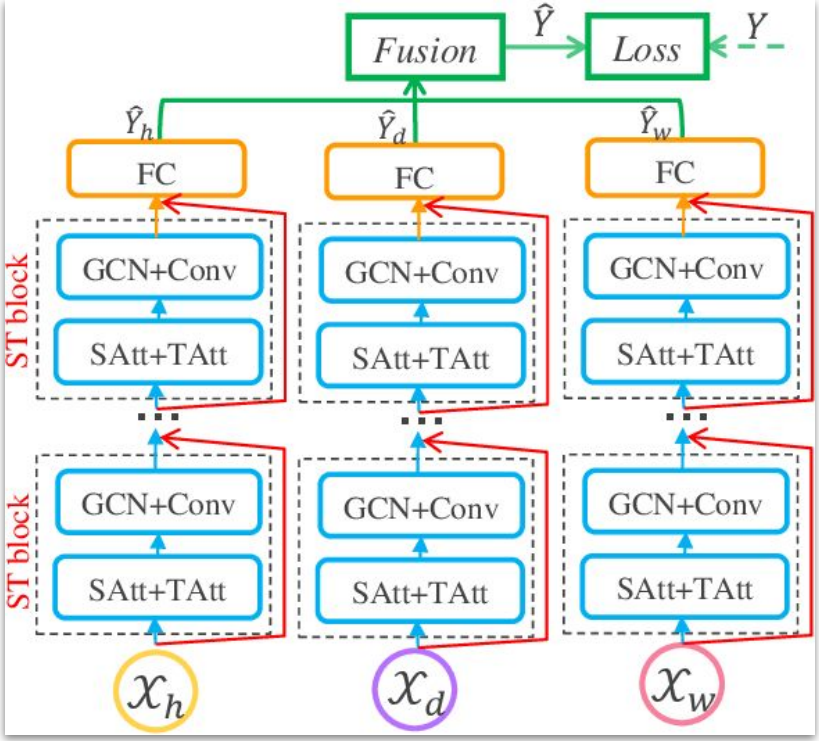
Graph Convolutional Networks

- Draws on the same idea as CNNs
  - Capture neighbourhood information for graph nodes
- Spatial
  - Aggregate neighbours feature information
- Spectral
  - Spectral graph analysis
  - Laplacian
    - $L = D - A$



Architecture

- Three components which are fused
  - Hourly
  - Daily
  - Weekly
- *Spatial-Temporal Blocks*
  - Spatial Attention + Temporal Attention
  - GCN + Conv
  - Fully Connected Layer
  - MSE as loss functions



Spatial- and Temporal Attention

Spatial attention mechanism

- Model complex spatial correlations between the different locations

$$\mathbf{S} = \mathbf{V}_s \cdot \sigma((\boldsymbol{\mathcal{X}}_h^{(r-1)}\mathbf{W}_1)\mathbf{W}_2(\mathbf{W}_3\boldsymbol{\mathcal{X}}_h^{(r-1)})^T + \mathbf{b}_s)$$

$$\mathbf{S}'_{i,j} = \frac{\exp(\mathbf{S}_{\mathrm{i,j}})}{\sum_{j=1}^{N}\exp(\mathbf{S}_{\mathrm{i,j}})}$$

$$\mathbf{V}_s, \mathbf{b}_s \in \mathbb{R}^{N\times N}, \mathbf{W}_1 \in \mathbb{R}^{T_{r-1}}, \mathbf{W}_2 \in \mathbb{R}^{C_{r-1}\times T_{r-1}}, \mathbf{W}_3 \in \mathbb{R}^{C_{r-1}}$$

Temporal attention mechanism

- Capture the dynamic temporal correlations between different time slices

$$\mathbf{E} = \mathbf{V}_e \cdot \sigma(((\boldsymbol{\mathcal{X}}_h^{(r-1)})^T\mathbf{U}_1)\mathbf{U}_2(\mathbf{U}_3\boldsymbol{\mathcal{X}}_h^{(r-1)}) + \mathbf{b}_e)$$

$$\mathbf{E}'_{i,j} = \frac{\exp(\mathbf{E}_{\mathrm{i,j}})}{\sum_{j=1}^{T_{r-1}}\exp(\mathbf{E}_{\mathrm{i,j}})}$$

where $\mathbf{V}_e, \mathbf{b}_e \in \mathbb{R}^{T_{r-1}\times T_{r-1}}, \mathbf{U}_1 \in \mathbb{R}^{N}, \mathbf{U}_2 \in \mathbb{R}^{C_{r-1}\times N}, \mathbf{U}_3 \in \mathbb{R}^{C_{r-1}}$

Spatial-Temporal Convolution

- **Graph Convolutions**
  - Capture neighbouring information for each node on the graph
- **Temporal Convolution**
  - Extract temporal dependencies from nearby time slices

In- and output

tures of all the nodes at time t. $\boldsymbol{\mathcal{X}} = (\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_\tau)^T \in \mathbb{R}^{N \times F \times \tau}$ denotes the value of all the features of all the nodes over τ time slices. In addition, we set $y_t^i = x_t^{f,i} \in \mathbb{R}$ to rep-

$$\hat{\mathbf{Y}} = \mathbf{W}_h \odot \hat{\mathbf{Y}}_h + \mathbf{W}_d \odot \hat{\mathbf{Y}}_d + \mathbf{W}_w \odot \hat{\mathbf{Y}}_w$$

$$MSE = \frac{1}{n} \Sigma \underbrace{\left(y - \widehat{y} \right)^2}_{\text{The square of the difference between actual and predicted}}$$

Dataset

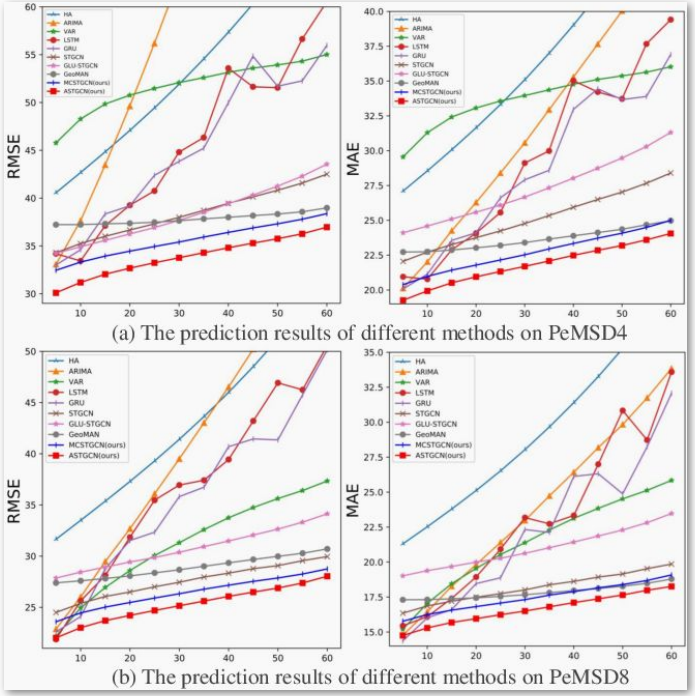
- Caltrans Performance Measurement System (PeMS)
- PeMSD4
  - 3848 detectors on 29 roads
  - Used **307** detectors
- PeMSD8
  - 1979 detectors on 8 roads
  - Used **170** detectors

[Image source]

Experimental setting and results

Model	PeMSD4		PeMSD8	
	RMSE	MAE	RMSE	MAE
HA	54.14	36.76	44.03	29.52
ARIMA	68.13	32.11	43.30	24.04
VAR	51.73	33.76	31.21	21.41
LSTM	45.82	29.45	36.96	23.18
GRU	45.11	28.65	35.95	22.20
STGCN	38.29	25.15	27.87	18.88
GLU-STGCN	38.41	27.28	30.78	20.99
GeoMAN	37.84	23.64	28.91	17.84
MSTGCN (ours)	**35.64**	**22.73**	**26.47**	**17.47**
ASTGCN (ours)	**32.82**	**21.80**	**25.27**	**16.63**

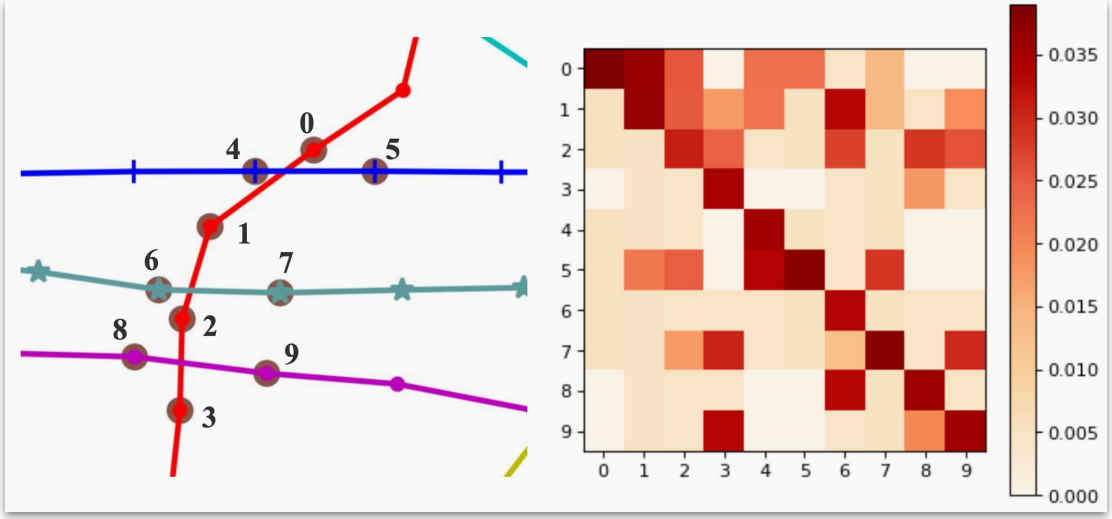
Table 1: Average performance comparison of different approaches on PeMSD4 and PeMSD8.



(a) The prediction results of different methods on PeMSD4

(b) The prediction results of different methods on PeMSD8

Interpretability



Conclusions and thoughts

- A novel *Attention Based Spatial-Temporal* GCN has been proposed
- Superior to existing models at the time

- I really enjoyed researching the topic
- Not relevant for my thesis directly
- Code is available on GitHub
  - MxNet implementation
  - PyTorch implementation

- Found Graph Neural Networks very interesting

Thanks!