Traffic Flow Forecasting During Covid-19 Pandemics

Group: Xinming Pan, Han Wang, Fangyang Zhang

Table of content

Motivation

Forecasting traffic flow during Covid-19 is crucial for following reasons

1. Covid-19 Containment
2. Resource Allocations
3. City Planning
4. Public Safety
5. Environmental Concerns

Dataset

The dataset is collected from traffic flow data in New York City from 2020-08-10 to 2022-4-25

- 90 weeks, 127 postal address, 5 features
- attribute of interest
  - placekey
  - total_population
  - median_income
  - case_rate
  - raw_visit_counts

Linear Regression

Use sklearn package and import linear regression

Build multivariate linear regression and fit the model, because here we are going to use four features to predict the number of visits (p=4).

The general formula for multivariate linear regression is as below:

$$y_i = b_0 + \sum_{j=1}^{p} b_j x_{ij} + e_i$$

Historical Average

- HA model calculate the average of a time series over a specific period, such as the past few years, and use that value to forecast future values.
- In our data set, we computed the number of visits by taking the average of visits in past several weeks.

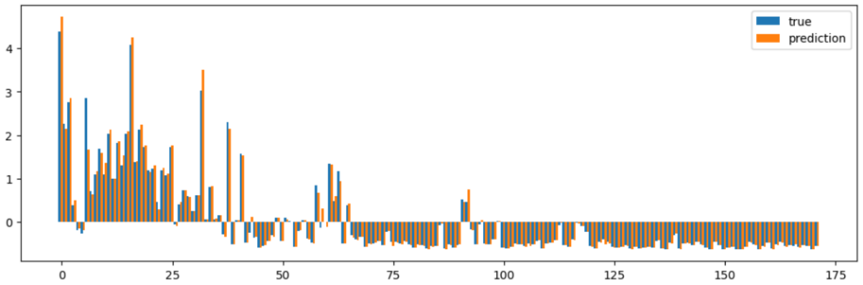
$$y_0 = x_0$$

$$y_t = \frac{x_t + x_{t-1} + x_{t-2} + \ldots + x_{s-t-1}}{s}$$

AR(p) Autoregressive Model

AR(p) model is a typical time series model, the main idea of this model is try to make a connection between past and present. The general formula for the AR(p) model is:

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

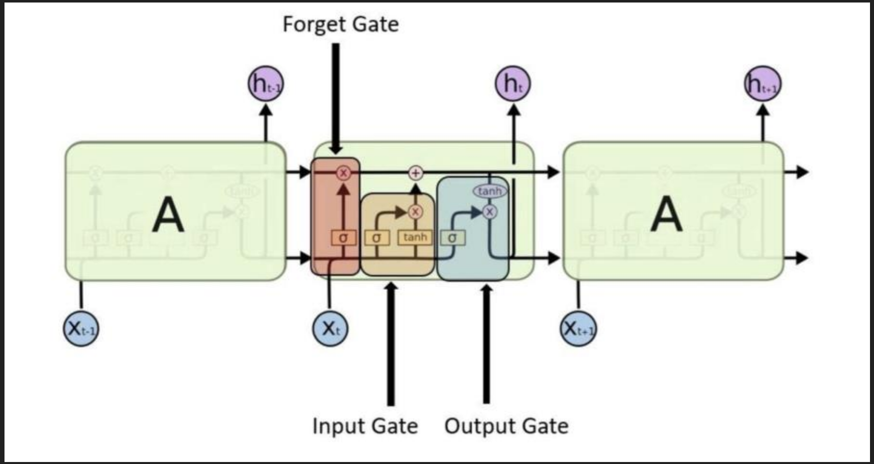


Here, we use 90 weeks data to build the model and try AIC test to find the most precise AR model to fix the data, and in addition, based on the PACF plot, we found that it is best in lags = 1, so we use AR(1) model in this data.

Long Short Term Memory

Long Short Term Memory (LSTM) is a type of Recurrent Neural Network that address the vanishing gradient problem of vanilla RNN.

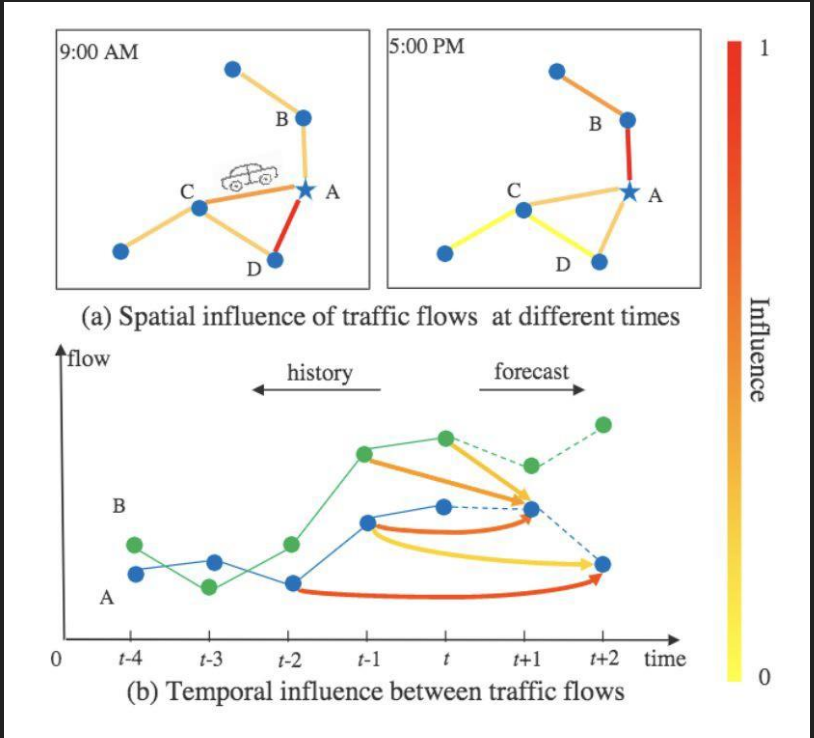
It consists of input gate, forget gate, and output gate which controls information flow between memory cells so it can remember information over long time periods



Attention based Spatial Temporal Graph Convolution Network (ASTGCN)

Traditional time series model doesn’t take into account the spatial correlation between traffic data

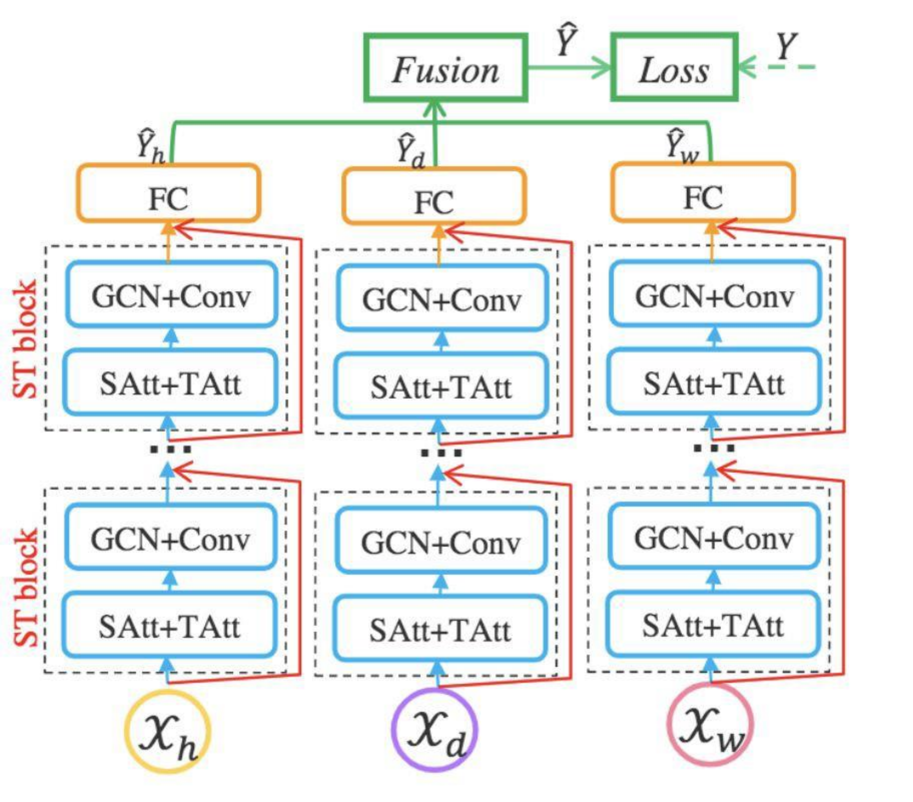
This spatial-temporal correlation in traffic data prompts us to apply spatial temporal attention in graph convolution model to solve the prediction task



(a) Spatial influence of traffic flows at different times
(b) Temporal influence between traffic flows

ASTGCN Architecture

- Spatial Attention
  - capture the dynamic correlations between nodes in the spatial dimension
- Temporal Attention
  - capture the temporal correlations between traffic conditions in different time slices
- Graph Convolution Network
  - traffic network is a graph in nature, and the features of each node can be regarded as signals of graph
  - use graph convolution to directly process each signals



Results

- Evaluation Criterion
  - Pearson Correlation of Coefficient
  - Mean Square Error
- Drawback of Pearson R
  - Two variables are not sampled independently from the distribution (bias in estimate)
  - Does not consider seasonal patterns or non-linear trends in time series data
  - Does not take into account the magnitude of the errors between the predicted and actual values

	HA	AR	Linear Regression	LSTM	ASTGCN
MSE	0.515	0.202	0.311	0.397	0.197
Pearson R	0.951	0.901	0.834	0.786	0.963

Results

- After leveraging spatial-temporal correlation, ASTGCN outperforms all the models in both Pearson R and MSE metrics