STGCN:SPATIAL-TEMPORAL GRAPH NETWORK FOR PANDEMIC PREDICTION

Ran Li
Data Science & Analytics, Information Hub
The Hong Kong University of Science and Technology
ran.li@connect.ust.hk

Botao Wang
Data Science & Analytics, Information Hub
The Hong Kong University of Science and Technology
bwangbk@connect.ust.hk

December 5, 2020

ABSTRACT

The COVID-19 pandemic prediction is important but also challenging. Classical transmission law and machine learning-based models can not utilize both the location and time-series information together to do the prediction. To solve this, we apply Spatial Temporal Graph network (STGCN) here. This novel network structure can capture the time-series information defined on a graph structure. By experiments, it is shown that this model perform well on the COVID-19 prediction tasks by successfully showing the trends of the pandemic.By choosing proper prediction window, model can focus on either long term or short term prediction.

***K*eywords** Pandemic prediction · deep learning · graph convolution

1 Introduction

In 2020, COVID-19, the new coronavirus spread all over the world and millions of people were infected. This 'global pandemic' situation has severely threatened public health and global economy. As reported, up to November 2020, more than 64.0 million people were affected and 1.5 million deaths were caused all over the world[1]. Predication of the pandemic can be essential for the government to make decision in advance to control the situation and make best use of the limited medical resources. However, COVID-19 prediction involves both spatial and temporal features which makes prediction challenging.

In the previous work, models in epidemiology and machine learning-based models are used to predict the pandemic situation. Different model has different limitations[2, 3]. Epidemiological models like Susceptible-infected-removed(SIR) model use transmission law based on the disease dynamic information. They are able to do a long-term prediction but will be bad at short-term prediction. Statistical model usually model each location separately, thus, neglect interaction between different node, meaning the geological information. A model that can capture both the geographical and temporal feature will be ideal to for the pandemic prediction. Inspired by the work using STGCN (Spatial-temporal graph convolutional networks) to do traffic forecasting[4], we decide to use this idea to do the COVID-19 prediction. For COVID-19 prediction, the most important thing is to predict the number of confirmed cases.The first problem is how to obtain the graph information to do the prediction. Here the node in the graph represent each county and for the edge, we consider both the geological distance and population of different places to define the weight. Spatial graph convolution is used to extract the graph feature and a Gated convolution is used to handle the times-series information[4]. The dataset about the number of confirmed cases and deaths in each state of the USA [5] are used to test our model. Results evaluated by MAPE shows that our model can successfully predict the trend of the pandemic.

2 Methods

2.1 Problem formulation

Pandemic prediction is a time series prediction task defined on graphs that using data in previous L_{in} time steps (L_{in} days) to predict the most likely confirmed cases in the next L_{out} time steps. The mathematical formulation is

$$\hat{c}_{t+1}, ..., \hat{c}_{t+L_{out}} = argmax_{c_{t+1},...,c_{t+L_{out}}} log\ P(c_{t+1}, ..., c_{t+L_{out}} | F_{t-L_{in}+1}, ..., F_t), \tag{1}$$

where c_t is the number of confirmed cases at time t for all the n nodes and F is the input with both number of confirmed cases and deaths.

2.1.1 Graph construction

To capture the spatial information, we generate the graph based on both the location information and population distribution. The graph structured data at time t is a attributed graph $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}, W)$.

$\mathcal{V}_t$ is the vertices. Each node represents a county in the United States associated with a feature vector containing the number of confirmed cases and deaths .

$\mathcal{E}$ is the edges and the connectivity is measured by the weight W. The new coronavirus can be passed from person to person, to measure the process, population density is important. The higher the populations density, the more likely for the virus to spread and this is also decide the population flow between two counties. Another factor is the geographical proximity which decides how likely it is for one person moving between the two counties. Based this assumption, the weighted adjacency matrix can be defined and the weight between node i and j is computed as

$$w_{ij} = p_i^{\alpha} p_j^{\beta} exp(-\frac{d_{ij}}{\delta^2}) \tag{2}$$

p_i is the population at node i (the county i) and d_{ij} is the physical distance between them and is computed based on the geological information.

2.2 Model

The model used is STGCN[4]. The components of the end to end model consists of two spatial-temporal convolutional blocks and followed by a fully-connected layer as the output. Within each spatial-temporal convolutional block, two temporal gated convolutional layer are connected by a spatial graph-convolution layer as a bridge.

Graph Convolution Our motivation is to analyze the spacial information to enhance the pandemic simulation. As we have learned in class, CNNs can capture the spatial feature of a image by moving the filters or kernels as a sliding window on the image. Usually, the same filters will be used in the same layer. This technique is workable here due to the grid nature of the image. In the pandemic situation. however, the counties distributed all around without a grid structure. To handle this, graph convolution technique is proposed[6]. There are two main GCN models, Spatial Graph Convolutional Networks and Spectral Graph and Convolutional Networks. Spatial one rearrange the nodes to a grid structure such that CNNs can handle. The spectral one used here apply graph Fourier transforms to solve the convolution in the spectral domains[6]. The idea is using message passing among different nodes to obtain the aggregated node features. The GCN can be represented as

$$Z = GCN(X, A) = \hat{D}^{\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} X \Theta \tag{3}$$

, where $X \in \mathbb{R}^{N \times C}$ is the feature of the N nodes and C dimensional feature. A is the adjacency matrix, $\hat{A} = A + I$, I is the identity matrix. Compared with A, $\hat{A}$ includes the self-loop, meaning the node is connected to itself. This makes message can pass to itself. $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$ is a diagonal matrix with each elements representing the degree of each node. Involving D acts as the role of normalization to the adjacency matrix. Θ is the parameters that need to be learned. The learned representation $Z \in \mathbb{R}^{N \times F}$ aggregate all the neighbor nodes feature and the weight is defined by the normalized adjacency matrix. Then a nonlinear transformation like ReLU will be acted on Z.

Gated CNNs Normally, recurrent neural networks with LSTM or GRU units are powerful tools for sequential data prediction. However, these models are slow to train since they process data sequentially while for CNN, all the data process simultaneously thus, much faster to train. It would be more efficient to apply CNN on sequential data, but in a normal CNN, a filter centring at the current time step will capture both the past and future information. To solve this problem, a Gated CNN with causal convolution is proposed[7]. A causal convolution only focus on the current and past

information by computing a normal convolution with input zero-padded with K-1 elements on the left. K is the kernel or filter size. The Gated CNN use causal convolution to obtain two convolution results A and B. The two results is then go through the gate, a gated linear unit(GLU) and process as $A \circ \delta(B)$, where $\circ$ stand for the element-wise hadamard product and δ is the sigmoid function. B decides what information of A can be passed to the next layer like a gate. The gate mechanism also provide non-linear transformation of the layer out put.

Spatial-temporal convolution block The Sandwich structure with two Gated CNN connected by the Spatial convolution in between is proposed in the traffic forecasting paper[4]. Firstly, pandemic prediction need both spatial and temporal information and this structure can combine the two. The input is then passed through the Gated CNN first to extract temporal information and then connect to Spatial Graph convolution to make use of the spatial feature defined on the graph. Finally the hidden state is passed back to a gated CNN. This can act like a bottleneck structure that first reduces the dimensionality of the input to process and then increase the dimensionality back., which is more computationally efficient.

3 Experiment

3.1 Dataset Description

In the experiment, we apply the STGCN to predict the number of COVID-19 confirmed cases in each state of the USA. The original data is available online[5], which includes the number of confirmed cases, the number of death cases in every county of different states from 21 Jan. until now. As the pandemic has not spread over all counties in the first few months where the statistical confirmed cases are sparse, the data only after 1st Jun. is used for the prediction in the experiment in the project. In order to apply the Data2Graph model learn the graph structure, the population density and geographical information (longitude and latitude) of each county are also used, which are also available online[5].

3.2 Data processing

For the date information, set the first day (1st Jun, 2020) as the start timestamp. The timestamp of other dates is defined as the difference from the first day. As we aims to predict the confirmed cases number only in the state perception, the summation of all confirmed cases number in each county is the study object of the given state. Besides, in order to study whether the death number would help to improve prediction of confirmed cases number, it is also selected as another dimension of input feature. Thus, two-dimension feature of the given state s and given timestamp t can be denoted as

$$F^t = [F^t_{s_1}, F^t_{s_2}, ..., F^t_S]^T, F^t_{s_i} = [F^t_{s_i,1}, F^t_{s_i,2}]^T \tag{4}$$

, where s_i represents the i-th state, S is the total state number which is a constant 52, $F^t_{s_i,1}$ is the confirmed cases and $F^t_{s_i,2}$ is the death number in i-th state. Then the data set D can be denoted as $D = F^t, t0, 1, 2, , T$ where T is the total timestamp number. The data with the timestamp $0 < t \leq 0.6T$,$0 < t \leq 0.6T$,$0.6T < t \leq 0.8T$,$0.8T < t \leq T$ serve as the training set, validation set and test set respectively. Given the length of input timestamp L_{in} and output timestamp L_{out}, the input and output of the GCN is

$$X = [F^{t_n}_s, F^{t_n+1}_s, ..., F^{t_n+L_in-1}_s], Y = [F^{t_n+L_in}_s, F^{t_n+L_in+1}_s, ..., F^{t_n+L_in+L_out-1}_s] \tag{5}$$

where the t_n is start timestamp. It is easy to see that if the timestamp of the given subset range in $[a, b]$, $a \leq t_n \leq b - L_{in} - L_{out}$, which means the sample number of this subset is $(b - a - L_{in} - L_{out})$.

3.3 Task and evaluation

Based on the county-level pandemic situation, we provide a state-level prediction by merging county data within the same state. By varying the length of prediction window and the number of time steps used to do the prediction, we test the model performance on relatively long-term and short-term predictions. The loss function of STGCN for pandemic prediction is,

$$L(\hat{c}; W) = \sum_t ||\hat{c}(c_{t-L_in+1}, ..., c_t, W) - c_{t+1}||^2 \tag{6}$$

The mean absolute percentage error (MAPE) is used to evaluate the model, which is defined as,

$$MAPE = \frac{1}{N} \sum_i^N |\frac{y_i - \hat{y_i}}{y_i}| \tag{7}$$

, where y_i is the label and $\hat{y_i}$ is the model prediction.

Table 1: MAPE

$(L_{in}\ L_{out})$	(15 2)	(15 4)	(15 6)	(15 8)
MAPE	0.11	0.12	0.16	0.14

4 Results

We have tested our model performance based on the different length of input data and prediction windows. Generally, our model can predict the trends for COVID-19 pandemic situation.

$$MAPE = 0.10$$

when $(L_{in}, L_{out}) = (13, 4)$, that is, the MAPE value for the prediction about the number of confirmed cases in two days based on the previous 13 days data is 0.10. In the experiment, the number of input timestamp range from 11 to 25 and the number of output timestamp range from 2 to 10, both of whose sample step is set to 2. Applying the proposed model to the given situations, the MAPEs can be obtained to measure the performance of the proposed model. The overall consequence is shown in Fig.1. Generally, with the decrease of input length and increase of the output length, the MAPE would become larger, meaning worse performance. The increase of output timestamp number would influence the accuracy of the proposed model much more, as the gradient on this axis is larger. It shows that the model's low efficiency for long term prediction. While, the stability of the variant input length is relevant higher.

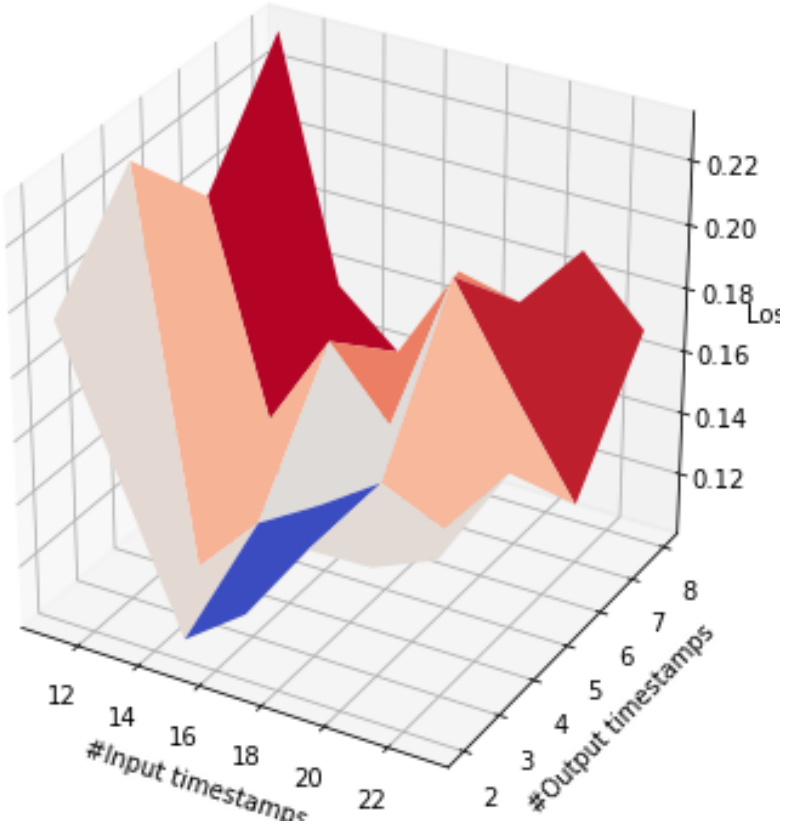
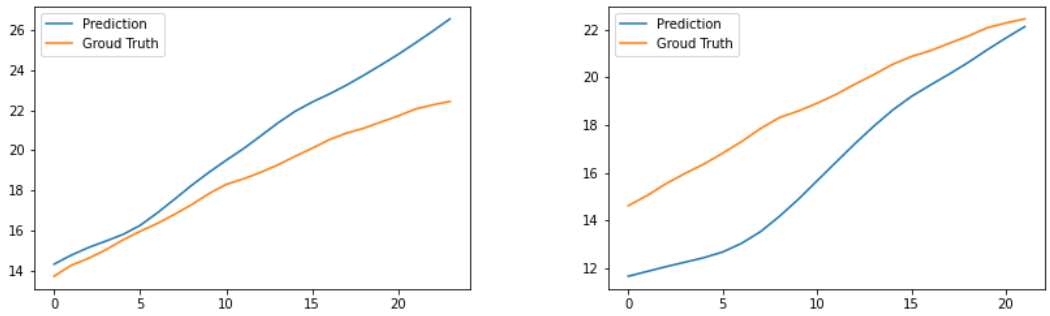


Figure 1: MAPE

With fixed length of input data, the performance decreases as the prediction window length increase (see table1). With chosen (L_{in}, L_{out}), the model can focus on short-term or long term prediction.

Given L_{in} days' confirmed case number as input, the output is the prediction of confirmed case number of 52 states in next L_{out} days. To illustrate the performance of the model, we take Maryland as the example, whose predicted



(a) $(L_{in}, L_{out}) = (11, 2)$ (b) $(L_{in}, L_{out}) = (11, 4)$

Figure 2: Trends

trends with different parameters are shown Figure 2. Figure 2 shows the different predicted trends, fig. 2(a) focus on short-term prediction, and fig. 2(b) perform better on the long-term prediction.

5 Discussion

The model can predict the trends but can not performs well on both long-term and short-term prediction.There are two possible reasons to explain. On the one hand, the feature we use is only about number of confirmed cases and deaths. The MAPE value of 0.10 is caused by not using the full information (see fig 3).For pandemic prediction, concept like active cases, removed cases are also important.

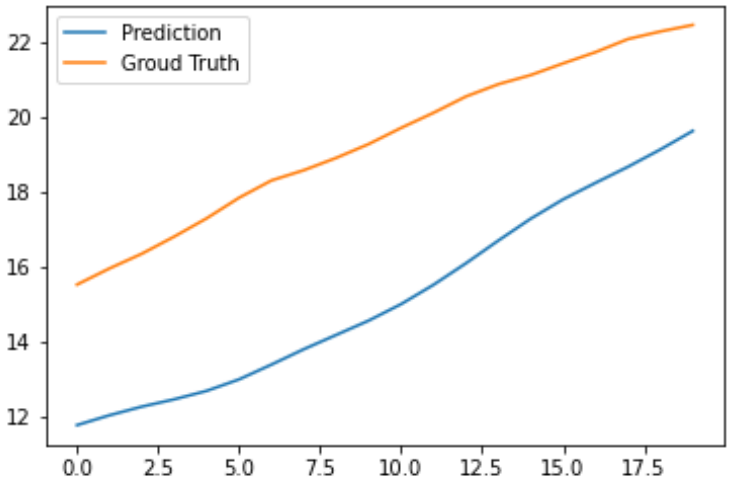


Figure 3: $(L_{in}, L_{out}) = (11, 6)$

On the other hand, while the model can extract both the graph feature and time-series feature, the transmission law information is missing.

6 Conclusion

In the project, we apply the Spatial-temporal graph convolution network for COVID-19 pandemic prediction.The novel network structure can capture both the spatial graph feature as well as the time-series information. It is shown that this model perform well on the COVID-19 prediction tasks by successfully showing the trends of the pandemic. In the future, we will try to use more complete information about COVID-19 and involve the transmission law to improve the prediction.

References

[1] Wikipedia. Covid-19 pandemic data. December 2020.

[2] Sen Pei and Jeffrey Shaman. Initial simulation of sars-cov2 spread and intervention effects in the continental us. *medRxiv*, 2020.

[3] Zifeng Yang, Zhiqi Zeng, Ke Wang, Sook-San Wong, Wenhua Liang, Mark Zanin, Peng Liu, Xudong Cao, Zhongqiang Gao, Zhitong Mai, et al. Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions. *Journal of Thoracic Disease*, 12(3):165, 2020.

[4] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*, 2017.

[5] The New York Times. Covid-19 data. December 2020.

[6] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[7] Yann N Dauphin, Angela Fan, Michael Auli, and David Grangier. Language modeling with gated convolutional networks. In *International conference on machine learning*, pages 933–941. PMLR, 2017.