A Spatial-temporal Graph based Hybrid Infectious Disease Model with Application to COVID-19

Yunling Zheng[1], Zhijian Li[1], Jack Xin[1] and Guofa Zhou[2]

[1]*Department of Mathematics, UC Irvine, U.S.A.*
[2]*Department of Health Sciences, UC Irvine, U.S.A.*

Keywords: COVID-19, Machine Learning, Spatial-Temporal, Graph RNN.

Abstract: As the COVID-19 pandemic evolves, reliable prediction plays an important role in policymaking. The classical infectious disease model SEIR (susceptible-exposed-infectious-recovered) is a compact yet simplistic temporal model. The data-driven machine learning models such as RNN (recurrent neural networks) can suffer in case of limited time series data such as COVID-19. In this paper, we combine SEIR and RNN on a graph structure to develop a hybrid spatio-temporal model to achieve both accuracy and efficiency in training and forecasting. We introduce two features on the graph structure: node feature (local temporal infection trend) and edge feature (geographic neighbor effect). For node feature, we derive a discrete recursion (called I-equation) from SEIR so that gradient descend method applies readily to its optimization. For edge feature, we design an RNN model to capture the neighboring effect and regularize the landscape of loss function so that local minima are effective and robust for prediction. The resulting hybrid model (called IeRNN) improves the prediction accuracy on state-level COVID-19 new case data from the US, out-performing standard temporal models (RNN, SEIR, and ARIMA) in 1-day and 7-day ahead forecasting. Our model accommodates various degrees of reopening and provides potential outcomes for policymakers.

1 INTRODUCTION

The classical infectious disease model, SEIR model (Hethcote, 2000), is a variation of the basic SIR model (Anderson and May, 1992). It assumes that all individuals in the population can be categorized into one of the four compartments: Susceptible, Exposed, Infected and Removed, during the period of pandemic. The model describes the evolution of the compartmental populations in time by a system of nonlinear ordinary differential equations (ODE):

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\beta_1 S I \\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \beta_1 S I - \sigma_1 E \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \sigma_1 E - \gamma I \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \gamma I
\end{aligned}
$$

The total population $S + E + I + R$ is invariant in time, which we shall normalize to 1 or 100 % in the rest of this paper. Clearly, SEIR is a simplistic temporal model of a given region or country.

However, the infectious disease data often provides not only temporal but also spatial information as in the case of COVID-19, see (Dong et al., 2020). A natural idea is to elevate SEIR model to a spatio-temporal model so that it can be trained from the currently reported data and make more accurate real-time prediction. See (Roosa et al., 2020) for temporal modeling on cumulative cases of China and real-time prediction.

In this paper, we set out to model the latent effect of inflow cases from the geographical neighbors to capture spacial spreading effect of infectious disease. For the practical reason that the inflow data is not observable, machine learning methods such as regression and neural network are more suitable. As widely adopted in time-series prediction problem, linear statistical models, such auto-regressive model (AR) and its variants are standard methods to forecast time-series data with some distribution assumptions on the time series. And the Long Short Term Memory neural networks model (LSTM) (Hochreiter and Schmidhuber, 1997) for the natural language processing problem, can be applied to time series data, especially disease data. With additional spatial information, the graph-structured LSTM models show a better performance on spatio-temporal data. See the application to

357

influenza data (Li et al., 2019)(Deng et al., 2019) and crime(Wang et al., 2019) and traffic data (Yu and Yin, 2018). However, such neural network models have a demand for a large training data to optimize the high dimensional parameters. Yet the reliable daily data of COVID-19 in the US begins after March 2020 and limits the temporal resolution. Applying space-time LSTM models (Li et al., 2019; Wang et al., 2018) directly to COVID-19 may lead to overfitting. In light of the shortage of data of COVID-19, we shall derive a hybrid SEIR-LSTM model with much fewer parameters than space-time LSTMs (Lai et al., 2017)(Wu et al., 2018).

2 RELATED WORK

In (Yang et al., 2015), ARGO (AutoRegression with Google search trends), a variant of AR, uses the google search trends to generate external feature of ARGO and forecasts influenza data from Centers for Disease Control of U.S.(CDC). ARGO is a linear statistical model that combines historical observations and external features. The prediction of influenza activity level is given by:

$$\hat{y}_t = u_t + \sum_{j=1}^{52} \alpha_j y_{t-j} + \sum_{i=1}^{100} \beta_i X_{i,t}.$$

where $\hat{y}_t$ is the predicted value at time t, and the optimization part of ARGO is:

$$\min_{\mu_y, \vec{\alpha}, \vec{\beta}} \left(y_t - u_t - \sum_{j=1}^{52} \alpha_j y_{t-j} - \sum_{i=1}^{100} \beta_i X_{i,t} \right)^2$$

$$+ \lambda_a ||\vec{\alpha}||_1 + \eta_a ||\vec{\beta}||_1 + \lambda_b ||\vec{\alpha}||_2^2 + \eta_b ||\vec{\beta}||_2^2$$

where $\vec{\alpha} = (\alpha_1, \cdots, \alpha_{52})$ and $\vec{\beta} = (\beta_1, \cdots, \beta_{100})$. y_{t-j} ($1 \leq j \leq 52$) are historical values of past 52 weeks and $X_{i,t}$ ($1 \leq i \leq 100$) are the google search trend features at time t. The feature are generated by top 100 of most related trends to influenza from google search at each time. The additional regularization terms to linear regression model helps ARGO optimize. The numerical experiment from (Yang et al., 2015) shows a better performance than machine learning models such as LSTM, AR, and ARIMA.

The (Li et al., 2019) introduces a graph structured recurrent neural network (GSRNN) to further improve the forecasting accuracy of CDC influenza activity level data. From CDC data, the USA is divided into 10 Health and Human Services (HHS) regions to report influenza activity level. These 10 regions are described as a graph in GSRNN with nodes $v_1, \cdots, v_{10}$

and a collection of edges based on geographic neighbor relationship (i.e. $E = \{(v_i, v_j)|v_i, v_j \text{ are adjacent}\}$, E is the set of all edges). By comparing the average record of activity levels, the 10 HHS region nodes are divided into two groups by relatively active level, the high active group $\mathcal{H}$, and low inactive group $\mathcal{L}$. The two group leads to 3 types of edges between them, $\mathcal{L} - \mathcal{L}$, $\mathcal{H} - \mathcal{L}$, and $\mathcal{H} - \mathcal{H}$, where each edge type has a customized RNN, called edge-RNN, to generate the edge features. There are also two kinds of RNNs for each node group to combine the edge feature with historical values and output the final prediction. Suppose a fixed node $v \in \mathcal{H}$. The edge feature of v at time t are $e_{v,\mathcal{H}}^t$ and $e_{v,\mathcal{L}}^t$, which are generated by the average of historical values of neighbor nodes of v in corresponding groups. The edge features are the input of the corresponding edge-RNN of each edge:

$$f_v^t = \text{edgeRNN}_{\mathcal{H}-L}(e_{v,L}^t), \; h_v^t = \text{edgeRNN}_{\mathcal{H}-\mathcal{H}}(e_{v,L}^t)$$

Then, the outputs of edge-RNNs are fed into the node-RNN of group $\mathcal{H}$ together with the node feature of v at time t, denoted as v^t, to output the prediction of the activity level of node v at time $t+1$, or y_v^{t+1}:

$$y_v^{t+1} = \text{nodeRNN}_{\mathcal{H}}(v^t, f_v^t, h_v^t).$$

3 OUR APPROACH: IeRNN MODEL

We propose a novel hybrid spatio-temporal model, named IeRNN, by combining LSTM (Hochreiter and Schmidhuber, 1997) and I-equation on a graph structure. The I-equation is a discrete in time model derived from SEIR differential equations. It resembles a nonlinear regression model of time series. The LSTM framework is applied to model the latent geographical inflow of infections. Our IeRNN model, comparing to (Li et al., 2019; Wang et al., 2019; Wang et al., 2018), is much more compact.

3.1 Derivation of I-equation from SEIR ODEs

As a variation to SEIR model, we shall construct additional features I_e and E_e that reveal the inflow population of infectious and exposed individuals from neighboring regions. Then we augment the SEIR differential equations with I_e and E_e as:

$$\frac{dS}{dt} = -\beta_1 SI - \beta_2 SI_e \tag{1}$$

$$\frac{dE}{dt} = \beta_1 SI + \beta_2 SI_e - \sigma_1 E - \sigma_2 E_e \tag{2}$$

$$\frac{dI}{dt} = \sigma_1 E + \sigma_2 E_e - \gamma I \tag{3}$$

$$\frac{dR}{dt} = \gamma I \tag{4}$$

It still follows that

$$S + E + I + R = 1 \tag{5}$$

by normalizing compartmentalized populations to percentages of total population. From (1) and (4), we have

$$R(t) = R(t_0) + \gamma \int_{t_0}^{t} I(\tau) d\tau \tag{6}$$

$$S = S_0 \exp\left(-\int_{t_0}^{t} (\beta_1 I + \beta_2 I_e) d\tau\right) \tag{7}$$

Substituting (6), (7) and (5) in (3), we have a closed I-equation:

$$\gamma I + \frac{dI}{dt} - \sigma_2 E_e = \sigma_1 \left(1 - I(t) - R(t_0) - \gamma \int_{t_0}^{t} I(\tau) d\tau \right.$$
$$\left. - S_0 \exp\left(-\int_{t_0}^{t} (\beta_1 I + \beta_2 I_e) d\tau\right)\right) \tag{8}$$

The above derivation holds for time dependent coefficients $\beta_i = \beta_i(t)$, $i = 1, 2$. Let $E_e = \tau I_e$, and write $\sigma_2 \tau$ as σ_2. By the explicit Euler and $(P+1)$-term Riemann sum approximation, we have a discrete time recursion:

$$\gamma I_t + I_{t+1} - I_t - \sigma_2 I_{e,t}$$
$$= \sigma_1 \alpha - \sigma_1 I_t - \gamma \frac{t - t_0}{P + 1} \sum_{j=0}^{P} I_{t-j}$$
$$- S_0 \exp\left(-\frac{t - t_0}{P + 1} \sum_{j=0}^{P} (\beta_1 I)_{t-j} + (\beta_2 I_e)_{t-j}\right) \tag{9}$$

which gives the *I-model*:

$$I_{t+1} = \sigma_1 \alpha + (1 - \sigma_1 - \gamma) I_t + \sigma_2 I_{e,t} - \gamma \frac{t - t_0}{P + 1} \sum_{j=0}^{P} I_{t-j}$$
$$- S_0 \exp\left(-\frac{t - t_0}{P + 1} \sum_{j=0}^{P} (\beta_1 I)_{t-j} + (\beta_2 I_e)_{t-j}\right) \tag{10}$$

If $I_e \equiv 0$ in I-model (10), we get an approximation of the I_t component of SEIR model, a nonlinear regression model in time for a single region, named the *I-equation*.

Since the official health agency, like CDC, did not track the migration of infectious and exposed cases nationwide, it is difficult to measure the affection from neighboring regions, here we model $I_{e,t}$ as a latent feature in absence of a mathematical formula or equation. To represent the latent feature from time-varying influx of infectious individuals, we make use of LSTM, a recurrent form of neural networks, see Fig. 1.
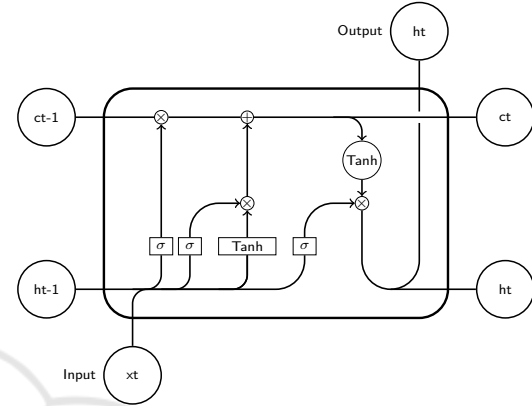

Figure 1: LSTM cell.

3.2 Generate Edge Feature with I_e

The spatial information based on US states map (Andrew, 2005), see Fig. 2, is formulated as an adjacent matrix $G = (g_{i,j})$. If two states v_i, v_j are neighbors to each other, then $g_{i,j} = 1$ otherwise is zero. With the variables of graph information, we can define the edge feature of state v_i at time t:

$$f_{i,t} = \frac{1}{\sum_j g_{i,j}} \sum_j \left(g_{i,j} \sum_{k=1}^{p} I_{j,t-k}\right)$$

where $I_{j,t}$ is the infectious population percentage in state v_j at time t.

Figure 2: USA state map.

Then we design an edge-RNN composed of stacked LSTM cells Fig. 3 with a following dense layer Fig. 4 to output I_e. The edge feature $f_{i,t}$ is the input of the edge-RNN. The integrated procedure to generate edge feature is illustrated by Fig. 5, taking California as example. Our model, IeRNN, is named by this design of edge-RNN for I_e and the I-equation:

$$I_{e,t} = \textit{Dense-Layer}(\textit{edge-RNN}(f_{i,t})) \qquad (11)$$
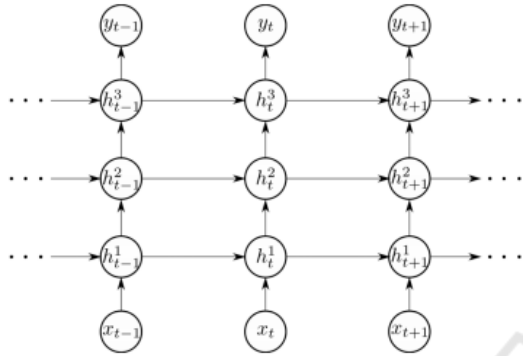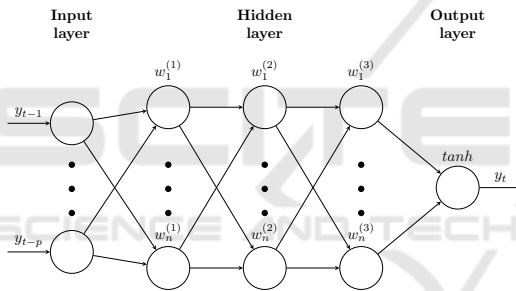

Figure 3: Stacked LSTM cells in edge-RNN.

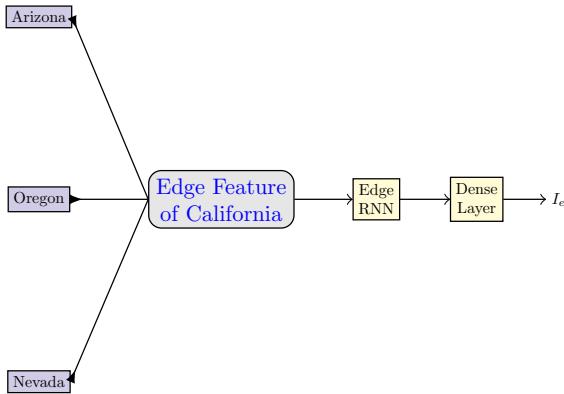Figure 4: Fully connected dense layer.

Figure 5: Generate I_e of California.

3.3 Policy Response Modeling

During an epidemic, the rate of infection could change as governments start responding to the epidemic. The infectious rate would start decreasing due

to the restrictive policy (partial or full lock-down) being put in place. We model the policy response by changing the parameter, β_1, in the ODE set (10). By multiplying a control factor β_{decay} (called test decay) to β_1, we can control the infection levels in the future resulting from various degrees of opening policies.

The policy response β_1 is a function of time (Li et al., 2020) to reflect no measure, restricting mass gatherings, reopening, lock-down for different times:

$$\begin{aligned}
\beta_1(t) = {} & \frac{2}{\pi}\arctan\left(\frac{-b(t-a)}{20}\right) \\
& + 1 + c\exp\left(-\frac{(t-t_0)^2}{s}\right)
\end{aligned} \qquad (12)$$

where parameters a, b, c, t_0, s are learned from fitting historical data.

4 EXPERIMENT

We use the COVID-19 data in United States(Dong et al., 2020) to evaluate our IeRNN model in training and testing. From (Dong et al., 2020), we find the state level infectious data in the US. Due to the incomplete recovered cases of US, we use the difference of cumulative cases in each state as the daily infected population. Then we use the population of US (World-Population-Review, 2020) to calculate the infectious rate of each state, where we assume the population of a state is constant during the period we concerned with. The data is split into training set (133 days) and testing set (35 days) for model evaluation.

The loss function for training is mean squared error (MSE) of the output of model and true data value:

$$\text{loss} = \frac{1}{T+1}\sum_{t=0}^{T}(I_t - \hat{I}_t)^2$$

where the output of model $\hat{y}_t$ has the form (adapted from (10)):

$$\begin{aligned}
\hat{I}_t = {} & \sigma_1\alpha + (1 - \sigma_1 - \gamma)I_{t-1} \\
& + \sigma_2 I_{e,t-1} - \gamma\frac{t-t_0}{P+1}\sum_{j=0}^{P}I_{t-1-j} \\
& - S_0\exp\left(-\frac{t-t_0}{P+1}\sum_{j=0}^{P}(\beta_1 I)_{t-1-j}\right. \\
& \left. \qquad\qquad + (\beta_2 I_e)_{t-1-j}\right)
\end{aligned} \qquad (13)$$

where we have parameters $(\alpha, \beta_1, \beta_2, \gamma, \sigma_1, \sigma_2)$. Due to the interpretations of SEIR model, these parameter values should range in the interval $[0,1]$.

We use gradient descent optimizer, Adam (Kingma and Ba, 2015), to train our IeRNN model. In each step, we update the weight of neural networks model (10) and the parameters of loss function (13) separately with different length of step and regularization norms.

To assess the performance of our model, we design a series of numerical experiments to compare the IeRNN with I-equation, temporal LSTM and ARIMA.

Regarding model size, the IeRNN and LSTM have about 4240 parameters while the I-equation and ARIMA have 5 parameters.

4.1 Robustness in Parameter Initialization

Model robustness in training is an important attribute, so that the model performance is not sensitive to initialization of parameters $(\alpha, \beta_1, \beta_2, \gamma, \sigma_1, \sigma_2)$ during training. We find that the I-equation (I-model with $I_e = 0$) is not easy to learn in the sense that a sub-optimal local minimum is often reached by gradient descent during optimization. With coupling to RNN ($I_e \neq 0$) in IeRNN, the landscape of loss function is regularized so that a local minimum from any random initialization gives a robust and accurate fit. Fig. 8 shows that I-equation is much less accurate in 1-day ahead prediction than IeRNN. Fig. 9 illustrates the same outcome in 7-day ahead prediction.

In further experiment, we train and test IeRNN and I-equation with randomly initialized parameters $(\alpha, \beta_1, \beta_2, \gamma, \sigma_1, \sigma_2)$. By repeating the training and testing procedure for 20 times, we compare the average MSE loss for both models. The results in Tables 1 and 2 show that IeRNN performs better for both training loss and testing loss in 1-day ahead and 7-day ahead predictions.

Table 1: Average MSE's of training (testing) loss in 1-day ahead prediction.

		IeRNN	I-equation
California	training	7.63e-09	8.49e-08
	testing	1.26e-08	9.68e-07
Florida	training	4.24e-08	3.45e-06
	testing	3.59e-08	3.97e-05
Virginia	training	3.70e-09	2.60e-08
	testing	6.90e-09	1.56e-07

4.2 1-day Ahead Prediction

We compare IeRNN (with $\beta_1(t)$), IeRNN, LSTM and ARIMA on 1-day ahead prediction. IeRNN achieves

Table 2: Average MSE's of training (testing) loss in 7-day ahead prediction.

		IeRNN	I-equation
California	training	8.01e-09	1.32e-07
	testing	9.66e-09	1.62e-06
Florida	training	8.15e-09	1.49e-06
	testing	9.77e-09	2.20e-05
Virginia	training	7.69e-09	8.21e-08
	testing	2.03e-08	1.41e-06

lower MSE error than LSTM and ARIMA on test set. With policy response function $\beta_1(t)$, IeRNN gives further improvement beyond IeRNN with constant β_1, see Table 3.

Table 3: MSE comparison of different models on 1-day ahead prediction.

	IeRNN $\beta_1(t)$	IeRNN	LSTM	ARIMA
California	1.83e-09	2.45e-09	5.00e-09	1.44e-08
Florida	6.13e-09	7.55e-09	4.68e-08	4.11e-08
Virginia	1.27e-09	1.29e-09	3.37e-09	3.74e-09

4.3 7-day Ahead Prediction

Motivated by weekly forecasting from CDC, we study the 7-day ahead prediction task. The loss function is modified by replacing the I-model by a 7-day delayed version below:

$$\hat{I}_t = \sigma_1 \alpha + (1 - \sigma_1 - \gamma) I_{t-7}$$
$$+ \sigma_2 I_{e,t-7} - \gamma \frac{t - t_0}{P + 1} \sum_{j=0}^{P} I_{t-7-j}$$
$$- S_0 \exp\left(-\frac{t - t_0}{P + 1} \sum_{j=0}^{P} (\beta_1 I)_{t-7-j}\right.$$
$$\left. + (\beta_2 I_e)_{t-7-j}\right)$$
$$(14)$$

where the output value at time t is influenced by the feature vector I_e at time $t - 7$ and earlier. With a similar modification of loss function, we adapt LSTM to the 7-day ahead prediction. Table 4 compares IeRNN and LSTM in terms of MSE on testing data.

4.4 Effect of Policy Response $\beta_1(t)$ in Testing

To study the effect of policy response in IeRNN model on testing data, we multiply the learned $\beta_1(t)$ by a constant factor (called *test decay*) during testing.
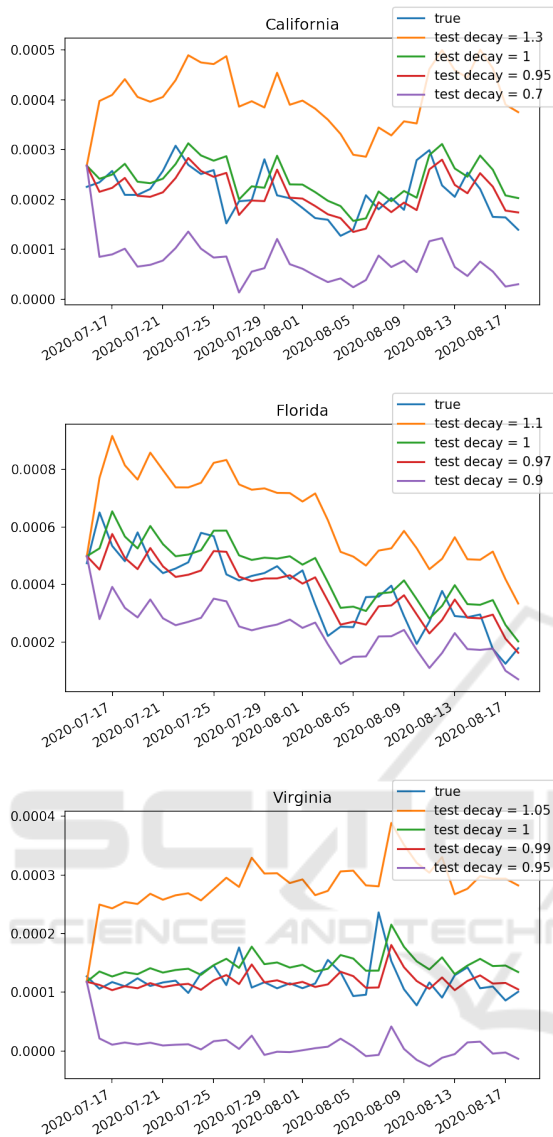
Figure 6: Effect of test decay (policy response multiplier) in test period of 1-day ahead prediction task. The IeRNN is trained through March 3, 2020 to July 14, 2020. The vertical axis is fraction of newly infected people in the population. The horizontal axis is time in unit of days.

Fig. 6 and Fig. 7 show the impact to model prediction on test data by adjusting *test decay* which could control the future trend of infection.

Table 4: MSE comparison of different models on 7-day ahead prediction.

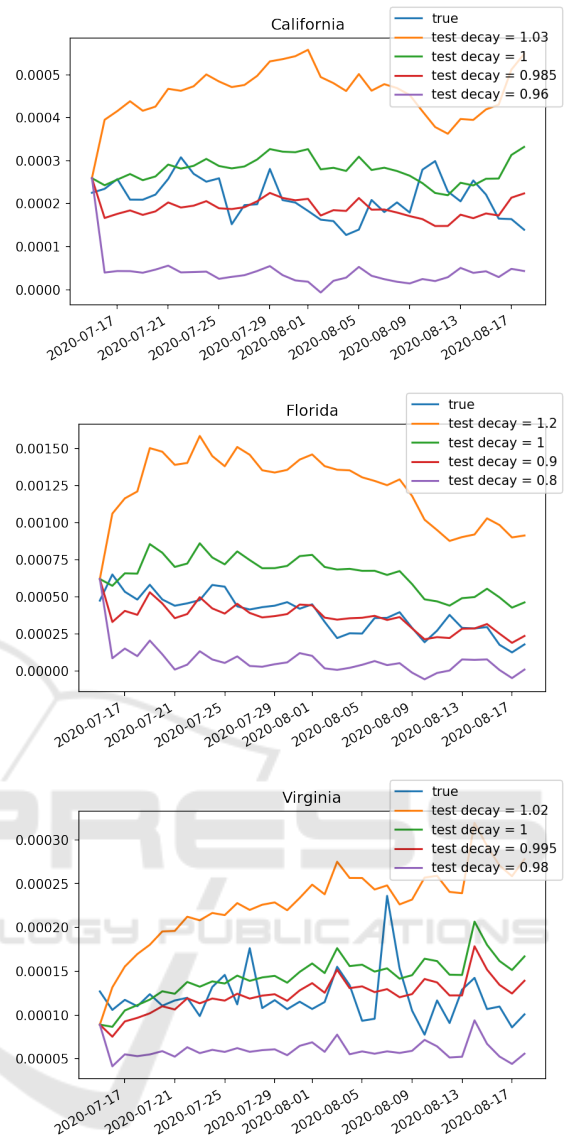|	IeRNN $\beta_1(t)$	IeRNN	LSTM
California	6.79e-09	9.84e-09	1.49e-08
Florida	4.34e-08	4.47e-08	5.74e-08
Virginia	1.16e-09	1.55e-09	1.54e-08

Figure 7: Effect of test decay (policy response multiplier) in test period of 7-day ahead prediction task. The IeRNN is trained through March 3, 2020 to July 14, 2020. The vertical axis is fraction of newly infected people in the population. The horizontal axis is time in unit of days.

5 CONCLUSIONS

We develop a novel spatio-temporal infectious disease model called IeRNN, which is a hybrid model consisting of I-equation from SEIR driven by spatial features. With such features and RNN dynamics as external input to the I-equation, the robustness to parameter initialization in model training is greatly improved. In 1-day and 7-day ahead prediction, our model outperforms standard temporal models. In future work,
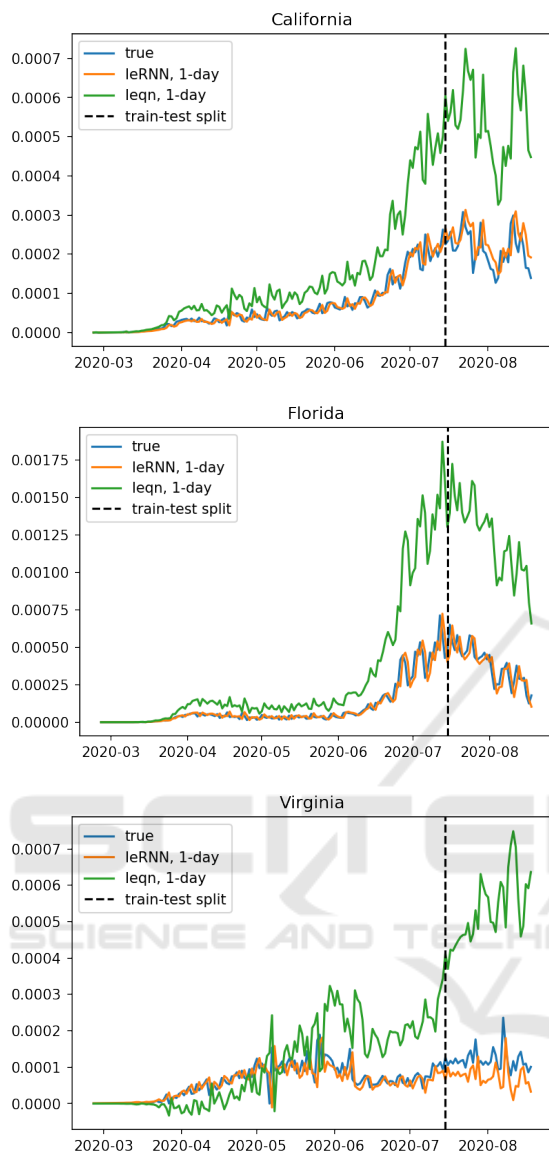
Figure 8: Comparing 1-day ahead predictions of IeRNN and I-equation with training (testing) period to the left (right) of the vertical dashed line. The vertical axis is fraction of newly infected people in the population. The horizontal axis is time in unit of days.
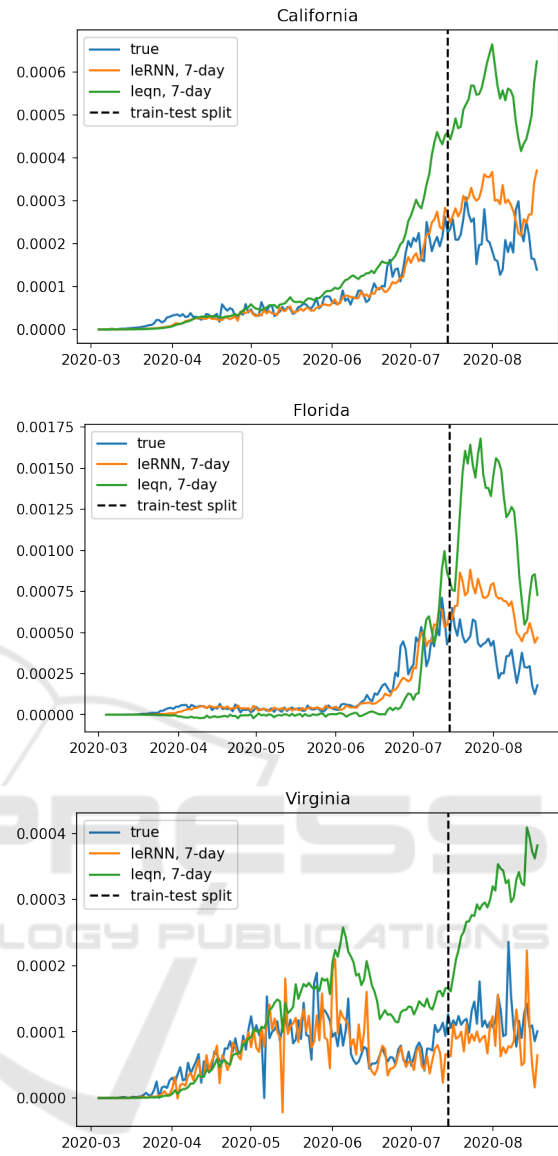
the social control mechanisms (Albi et al., 2020; Morris et al., 2020) could be considered to strengthen the I-equation, as well as traffic data to expand inflow effect beyond geographic neighbors.

ACKNOWLEDGEMENTS

Figure 9: Comparing 7-day ahead predictions of IeRNN and I-equation with training (testing) period to the left (right) of the vertical dashed line. The vertical axis is fraction of newly infected people in the population. The horizontal axis is time in unit of days.

REFERENCES

Albi, G., Pareschi, L., and Zanella, M. (2020). Control with uncertain data of socially structured compartmental epidemic models. *arXiv preprint arXiv:2004.13067*.

Anderson, R. and May, R. (1992). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford.

Andrew, C. (2005). A map of the united states, with state names (and washington d.c.).

Deng, S., Wang, S., Rangwala, H., Wang, L., and Ning, Y.

(2019). Graph message passing with cross-location attentions for long-term ili prediction. *arXiv preprint arXiv:1912.10202*.

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track covid-19 in real time. *Lancet Inf Dis. 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1*.

Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Review*, 42:599 – 653.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Kingma, D. and Ba, J. (2015). Adam: A method for stochastic optimization. *3rd International Conference for Learning Representations, San Diego, 2015*.

Lai, G., Chang, W., Yang, Y., and Liu, H. (2017). Modeling long- and short-term temporal patterns with deep neural networks. *CoRR*, abs/1703.07015.

Li, M. L., Tazi Bouardi, H., Skali Lami, O., Trikalinos, T. A., Trichakis, N. K., and Bertsimas, D. (2020). Forecasting covid-19 and analyzing the effect of government interventions. *medRxiv*.

Li, Z., Luo, X., Wang, B., Bertozzi, A., and Xin, J. (2019). A study on graph-structured recurrent neural networks and sparsification with application to epidemic forecasting. In *World Congress on Global Optimization*, pages 730–739. Springer.

Morris, D. H., Rossine, F. W., Plotkin, J. B., and Levin, S. A. (2020). Optimal, near-optimal, and robust epidemic control. *arXiv preprint arXiv:2004.02209*.

Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J., Yan, P., and Chowell, G. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, 5:256 – 263.

Wang, B., Luo, X., Zhang, F., Yuan, B., Bertozzi, A., and Brantingham, P. (2018). Graph-based deep modeling and real time forecasting of sparse spatio-temporal data. *MiLeTS '18, London, UK, DOI: 10.475/123_4; arXiv preprint arXiv:1804.00684*.

Wang, B., Yin, P., Bertozzi, A., Brantingham, P., Osher, S., and Xin, J. (2019). Deep learning for real-time crime forecasting and its ternarization. *Chinese Annals of Mathematics, Series B*, 40(6):949–966.

World-Population-Review (2020). Us states population 2020.

Wu, Y., Yang, Y., Nishiura, H., and Saitoh, M. (2018). Deep learning for epidemiological predictions. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*.

Yang, S., Santillana, M., and Kou, S. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO. *Proceedings of the National Academy of Sciences*, 112(47):14473–14478.

Yu, B. and Yin, H. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *Twenty-Seventh International Joint Conference on Artificial Intelligence IJCAI-18*.