Journal of Applied Analysis and Computation
Volume 12, Number 3, June 2022, 1077–1089
Website:http://www.jaac-online.com
DOI:10.11948/20220136

ESTIMATING TIME-VARYING REPRODUCTION NUMBER BY DEEP LEARNING TECHNIQUES*

Pengfei Song[1] and Yanni Xiao[1,†]

Dedicated to Professor Jibin Li on the occasion of his 80th birthday.

Abstract Estimating time-varying reproduction number $\mathcal{R}_t$ is important for quantifying the transmission ability, capturing the trend of infectious disease and assessing the effectiveness of public health intervention measures. However, accurate estimation of $\mathcal{R}_t$ remains a challenging work. Deep neural networks are uniform approximators and have an unreasonable and counter-intuitive effectiveness in learning unknown functions, thus can be applied to represent $\mathcal{R}_t$. In this paper, we will estimate $\mathcal{R}_t$ by universal differential equation method which embeds neural network $\mathcal{R}_t$ into a differential equation. Compared with other methods such as state space, EpiEstim and EpiNow2 methods, deep learning method can achieve better performance with fewer data sources.

Keywords Effective reproduction number, neural differential equations, neural networks, epidemic control.

MSC(2010) 92B20, 34H05, 65L99, 65P99.

1. Introduction

The basic reproduction number $\mathcal{R}_0$, defined as the expected average number of secondary cases produced in a completely susceptible population by a typical infectious individual during the infectious period, is one of the most significant concepts in population biology and epidemiology ([2, 18]). $\mathcal{R}_0$ determines the threshold behavior for many epidemic models and commonly used to estimate herd immunity level. It is often the case that a disease dies out if the basic reproduction number is less than unity and the disease is established in the population if it is greater than unity. The effective reproduction number, denoted as $\mathcal{R}_e$ or $\mathcal{R}_t$, is the expected average number of secondary cases produced by an infectious individual in a population where some individuals may no longer be susceptible. $\mathcal{R}_t$ often determines the potential for epidemic spread at a specific time t under the control measures. Estimation of $\mathcal{R}_t$ is crucial to assessing the effect of public health measures and

---

[†] The corresponding author. Email: yxiao@mail.xjtu.edu.cn(Y. Xiao)
[1] Department of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an City, 710049, China
[*] The authors were supported by National Natural Science Foundation of China (NSFC, 11631012(YX), 12101487(PS)) and China Postdoctoral Science Foundation (2020M683445(PS)).

vaccination effectiveness [28]. The effective reproduction number can also be used to forecast the turning point and trend of epidemic transmission.

Many effective reproduction number estimation methods arise these years ([23]), such as compartment models method ([19]), Wallinga-Teunis method ([42]), EpiEstim method ([12]), EpiNow2 method ([1]), EpiFilter method ([34]). However, these methods either can not achieve good performance with few data sources (case data, serial interval) or get accurate $\mathcal{R}_t$ at a cost of multiple data sources ([1]) such as case data, generation time distribution, incubation time distribution, reporting delays. Is there a way to mine enough information from fewer data and achieve better performance on estimating $\mathcal{R}_t$? Deep learning method can be an answer.

In recent years, deep neural network ([22, 29]), as a universal approximator for unknown mappings ([25]), shows an unreasonable effectiveness ([39]) in pattern recognition, learning unknown mechanisms and solving traditional difficult problems such as image recognition ([24]), natural language processing ([17]). Deep learning methods are also used to aid mathematicians in discovering new conjectures and theorems ([15]), and physicians in finding new physical laws ([13]). To better understand and improve the performance of deep neural networks on scientific problems such as physics, chemistry and epidemiology, more and more works pay attention to coupling or embedding differential equations and deep neural networks. One important idea is regarding deep neural network (DNN) as discretization of differential equation, which inspires researchers to redesign traditional sequential neural architectures based on numerical discretization schemes ([8, 10, 30, 33, 38]) or to replace DNN by a continuous model characterized by differential equations ([9, 16, 37]). Another revolutionary idea is training DNN from the view of optimal control. This extend the idea of backprogation ([22]) to include adjoint sensitivity analysis ([7]), which opens the mind in differential programming ([3]) and inspires a lot of architectures coupling neural network and differential equation, such as neural differential equation ([9]) and universal differential equations which embed neural networks into the differential equations ([36]).

Despite the remarkable progress in coupling the differential equation and deep learning, deep learning methods like universal differential equations ([36]) haven't attracted enough attention and widely used in epidemiological dynamics problems such as the expression of incidence rate, estimation of effective reproduction number. This may be because coupling of transmission models and deep learning requires the researchers to have a strong background of mathematical epidemiology and deep learning as well as a basic understanding of differential programming and scientific computation, which might not always be expected in aforementioned fields. To fill this gap, we apply deep learning method to estimate effective reproduction number, and investigate the advantages and weakness of deep learning method, which will also give new insights to traditional mathematical epidemiological problems.

In this paper, we will represent effective reproduction number $\mathcal{R}_t$ as

$$\mathcal{R}_t = \mathrm{NeuralNetwork}_\theta(t, I), \tag{1.1}$$

where $\mathrm{NeuralNetwork}_\theta(t, I)$ is a deep neural network receiving t, I as inputs and θ denotes the parameter set of neural network. Neural network is embedded into the following differential equation model (Fig. 1):

$$\begin{cases} I' = \gamma \mathrm{NeuralNetwork}_\theta(t, I)I - \gamma I, \\ H' = \gamma \mathrm{NeuralNetwork}_\theta(t, I)I, \end{cases} \tag{1.2}$$

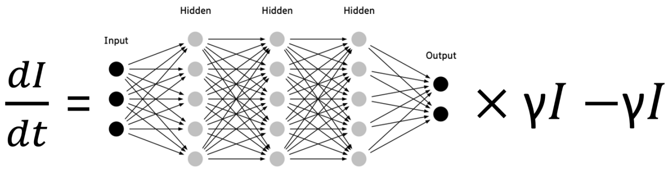

Figure 1. Scheme of deep neural network method.

where $I(t)$ and $H(t)$ denote the number of infected individuals and accumulated confirmed cases at time t, and $1/\gamma$ is infection period of infectious disease. Assumed that $\mathcal{T} = \{t_1, t_2, \cdots, t_n\}$ is a set of the case data sampling time points and $\mathcal{H}_o = \{h_1, h_2, \cdots, h_n\}$ denotes observed accumulated confirmed cases set. Then training of model (1.2) is defined as the following optimal problem:

$$\begin{aligned} &\min_{\theta} \mathrm{Loss}_{\theta}(\mathcal{H}_o, H(\mathcal{T})), \\ &\text{s.t.} \quad I' = \gamma \mathrm{NeuralNetwork}_{\theta}(t, I)I - \gamma I, \end{aligned} \tag{1.3}$$

where $H(\mathcal{T}) = \{H(t_1), H(t_2), \cdots, H(t_n))\}$ and $\mathrm{Loss}_{\theta}(\mathcal{H}_o, H(\mathcal{T}))$ is some kind of metric between $\mathcal{H}_o, H(\mathcal{T})$ such as log or square loss.

The rest of the paper is organized as follows. Universal differential equation method and adjoint sensitivity analysis will be introduced in Section 2. We will estimate $\mathcal{R}_t$ based on generated data and real COVID-19 case data by deep learning method, and compare deep learning method with three other $\mathcal{R}_t$ estimation methods in Section 3. Conclusion and discussion will be in Section 4.

2. Universal Differential Equation Method

In this part, we will introduce a recently proposed state-of-the-art methodology, universal differential equations (UDEs) ([36]), which couple differential equation models with deep learning structures for scientifically-based learning, and can be accurately extrapolate beyond the original data, obtain differentiable and fully observed fitting data, and accelerate model simulation, all in a time and data-efficient manner.

2.1. Deep Neural Network

To start with, we briefly introduce neural networks and the universal approximate properties. A deep neural network is a special kind of compositional function which applies linear and nonlinear transformations to the inputs recursively (See also a diagram of a deep neural network in Figure 1). Many deep neural networks architectures are proposed in recent years ([22]), such as feed-forward neural network (FNN), convolution neural network (CNN), recurrent neural network (RNN), long short-term memory (LSTM), gate recurrent unit (GRU). In this paper, the effective reproduction number $\mathcal{R}_t$ is represented by the simplest neural network architectures FNN (also called multi-layer perceptron (MLP)), which is sufficient for modeling emerging infectious disease. However, the ideas can also be extended to other neural network architecture. For example, CNN can be used to estimate effective reproduction number of spatial epidemic models.

Let $\mathcal{F}^L(x) : \mathbb{R}^{d_{in}} \to \mathbb{R}^{d_{out}}$ be an L-layer feed-forward neural network with an $(L-1)$-hidden layer neural network. $N_k, k = 0, 1, 2, \cdots, L$ denotes the number of neurons in the kth layer with $N_0 = d_{in}$ and $N_L = d_{out}$. Denote the weight matrix and bias vector in the kth layer by $W^k \in \mathbb{R}^{N_k \times N_{k-1}}$ and $b^k \in \mathbb{R}^{N_k}$, respectively. Then the FNN can be recursively defined as follows:

- input layer: $\mathcal{F}^0(\mathbf{U}) = \mathbf{U} \in \mathbb{R}^{d_{\text{in}}}$
- hidden layers: $\mathcal{F}^k(\mathbf{U}) = \sigma\left(W^k \mathcal{F}^{k-1}(\mathbf{U}) + b^k\right) \in \mathbb{R}^{N_k}$ for $1 \le k \le L-1$
- output layer: $\mathcal{F}^L(\mathbf{U}) = W^L \mathcal{F}^{L-1}(\mathbf{U}) + b^L \in \mathbb{R}^{d_{out}}$,

where W^k is a matrix or tensor, and σ is a nonlinear activation function. Commonly used activation functions include the logistic sigmoid $1/(1+\exp(-x))$, the hyperbolic tangent (tanh), the rectified linear unit (ReLU, $\max\{x, 0\}$) and swish function ($x*\text{sigmoid}(x)$) ([20]). Here, activation functions in different layers can be different.

One fundamental property of DNN is that a single layer neural network can simultaneously and uniformly approximate a function and its partial derivatives. Let $\mathbb{Z}^{d+}$ be d-dimensional nonnegative integers set. For $m = (m_1, \cdots, m_d) \in \mathbb{Z}^{d+}$, set $|m| := m_1 + \cdots + m_d$, and

$$D^{\mathbf{m}} := \frac{\partial^{|\mathbf{m}|}}{\partial x_1^{m_1} \ldots \partial x_d^{m_d}}.$$

We say $f \in C^m(\mathbb{R}^d)$ if $D^k f \in C(\mathbb{R}^d)$ for all $k \le m, k \in \mathbb{Z}^{d+}$, where $C(\mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{R} | f \textit{ is continuous}\}$ is the space of continuous functions. Then we have the following universal approximate theorem of derivatives using single hidden layer neural networks([35]).

Theorem 2.1 ([35]). *Let $m_i \in \mathbb{Z}^{d+}, i = 1, \cdots, s$, and set $m = \max_{i=1,\cdots,s} |m_i|$. Assume $\sigma \in C(\mathbb{R}^d)$ and that σ is not a polynomial. Then a single hidden layer neural network:*

$$\mathcal{M}(\sigma) := \mathbf{span}\left\{\sigma(\mathbf{w} \cdot \mathbf{U} + b) : \mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R}\right\}$$

is dense in

$$C^{\mathbf{m}^1, \ldots, \mathbf{m}^s}\left(\mathbb{R}^d\right) := \cap_{i=1}^s C^{\mathbf{m}^t}\left(\mathbb{R}^d\right).$$

Universal approximate theorems theoretically ensure that deep neural networks is able to overcome the inaccuracies of approximate unknown mappings or operators by directly learning the entire set of nonlinear interactions from data. In practice, deep neural networks method is shown to easily find good surrogate to validate the data. More importantly, good surrogate doesn't need to be unique, which implies a belief in practice that for large neural networks, a local minima is enough and global minima often leads to over-fitting, which need more theoretical investigating.

2.2. Universal Differential Equation

Before the introduction on UDEs, we introduce neural differential equations ([9]) first. Neural differential equations are initial value problems with the following form:

$$u' = \text{NN}_\theta(u, t), \tag{2.1}$$

where NN is a deep neural network receiving $[u, t]$ as input. Neural differential equations were proposed before UDEs, inspired the ideas in UDEs and can be regarded as a special case of UDEs. Neural differential equations make use of scientific structures as a modeling basis. Since the embedded neural network is a universal approximator, it follows that neural differential equations can learn to approximate any sufficiently regular differential equation. From the perspective of deep learning, neural differential equations are redisigned sequential neural networks based on numerical schemes of differential equation, and they can be regarded as continuous-depth or ”infinitely deep” ResNet-like deep learning models.

However, the resulting neural differential model is defined without direct incorporation of known mechanisms. UDEs extends the previous data-driven neural ODE approaches to directly utilize mechanistic modeling simultaneously with universal approximators. UDEs are initial value problems with the following forms:

$$u' = f_{\theta_2}(u, t, NN_{\theta_1}(u, t)), \tag{2.2}$$

where f is a known mechanistic model and NN denotes the missing or unknown terms, θ_1 and θ_2 are parameters of known mechanisms and neural networks, respectively, which can be estimated simultaneously. UDEs has stronger explaniablility than deep neural networks or neural differential equations, since they keep known mechanisms in physics, chemistry or epidemiology. UDEs are proved to be methods with good generalization and can be trained with less sample data ([36]).

2.3. Adjoint Sensitivity Analysis

The key issue of training universal differential equations is how to implement “backpropagation” ([22]) in differential equations. Actually, it is a mature topic in optimal control theory. From the perspective of optimal control, the training of neural differential equations can be regarded as optimal control problems, which extend the idea of backprogation ([22]) in differential programming to include adjoint sensitivity analysis ([7]).

In this part, we will introduce adjoint sensitivity analysis in detail. Considering the following optimal control problem:

$$\min_u \int_0^T g(u(t))dt, \quad \text{s.t.} \quad u' = f(u, \theta). \tag{2.3}$$

Here u can be regarded as neural networks or dynamics including neural networks. Our purpose here is to derive the ”backprogation” of derivatives of $\int_0^T g(u(t))dt$ to θ. If the loss function is discrete, i.e.,

$$\sum_{i=1}^n m(u(t_i)),$$

we can rewrite discrete loss function by Dirac function

$$\sum_{i=1}^n m(u(t_i)) = \int_0^T \sum_{i=1}^n m(u(t_i))\delta(t - t_i)dt.$$

Denote

$$L(\theta) = \int_0^T g(u)dt = \int_0^T g(u)dt + \int_0^T \lambda(t)\left(f(u, \theta) - u'(t)\right)dt.$$

Then

$$\begin{aligned}\frac{\partial L}{\partial \theta} &= \int_0^T g_u u_\theta dt + \lambda(T) u_\theta|_{t=T} + \int_0^T \lambda'(t) u_\theta dt \\ &\quad + \int_0^T \lambda(t)\left(f_u u_\theta + f_\theta\right) dt \\ &= \int_0^T \left[\lambda'(t) + g_u + f_u \lambda(t)\right] u_\theta dt + \int_0^T \lambda(t) f_\theta dt + \lambda(T) u_\theta|_{t=T}.\end{aligned}$$

Let

$$\lambda'(t) + g_u + f_u \lambda(t) = 0, \lambda(T) = 0. \tag{2.4}$$

We obtain

$$\frac{\partial L}{\partial \theta} = \int_0^T \lambda(t) f_\theta dt. \tag{2.5}$$

If the loss function is discrete, i.e.,

$$g(t) = \sum_{i=1}^{n} m(u(t_i))\delta(t - t_i),$$

then system (2.4) becomes an impulsive system

$$\begin{cases}\lambda'(t) = -f_u \lambda(t), \\ \lambda(t_i^+) = \lambda(t_i^-) + g_u(u(t_i)), i = 1, 2, \cdots, n.\end{cases} \tag{2.6}$$

3. Effective Reproduction Number by UDE Method

In this section, we estimate effective reproduction number $\mathcal{R}_t$ by UDE Method based on data generated by three toy infectious disease models and Ontario's first wave COVID-19 case data. The results show that UDE can fit the data well and deep neural networks represent $\mathcal{R}_t$ very well. Moreover, we compare UDE method with three other most used estimation methods: state space method, EpiEstim ([12]) method and EpiNow2 ([1]) method, and find that UDE method can achieve correct and smooth estimation with fewer data sources.

Deep learning methods are implemented in open source Julia language 1.6.2 ([5]). State space, EpiEstim ([12]) and EpiNow2 ([1]) methods are implemented in open source R language 4.1.2. All algorithms and codes in this section are available in https://github.com/Song921012/EstmatingRtDeepL

3.1. Toy Models

In this part, we generate data from the following three toy infectious disease models which have a lot of applications in epidemiology:

- subexpotential model ([31])

$$\begin{cases} I' = 0.2\sqrt{I}, \\ \mathcal{R}_t = \dfrac{2}{\sqrt{I}} + 1, \end{cases} \tag{3.1}$$

- logistic growth model

$$\begin{cases} I' = 0.2\left(1 - \dfrac{I}{30}\right) I, \\ \mathcal{R}_t = 3 - \dfrac{I}{15}, \end{cases} \tag{3.2}$$

- media impact model ([40])

$$\begin{cases} I' = 0.2\exp(-0.2I)\left(1 - \dfrac{I}{30}\right) I, \\ \mathcal{R}_t = 2\exp(-0.2I)\left(1 - \dfrac{I}{30}\right) + 1, \end{cases} \tag{3.3}$$

and compare real $\mathcal{R}_t$ in models (3.1)-(3.3) with estimated $\mathcal{R}_t = \text{NeuralNetwork}(t)$ in UDE (1.2) by deep learning method. Throughout this section, the infection period is assumed to be 10 days.

For subexpotential and logistic models, we use 30 days training data and make prediction for another 10 days. In media impact disease model, 20 days data is used for training and another 20 days data is used for prediction. The left and right parts in Fig 2 describe the fitting and prediction results, and $\mathcal{R}_t$ estimation results, respectively. From Fig. 2, we can see that universal differential equation model (1.2) can fit the generated data and predict well. Moreover, trained neural networks $\mathcal{R}_t$ match real $\mathcal{R}_t$ well, which gives insight to universal approximate abilities of deep neural networks.

3.2. Real COVID-19 Data

The UDE method is also used in estimating the effective reproduction number concerned with the first wave Covid-19 cases in Ontario, Canada, i.e., using UDE model (1.2) to fit the case data and training the neural networks $\mathcal{R}_t$. First wave data began on Feb 25, 2020 and continued for about 150 days. The accumulated and daily reported cases data are shown in Figs. 3(left) and 4(left), respectively. From Fig. 3(left) and 4(left), we can see that UDE model (1.2) fits first wave data well. $\mathcal{R}_t$ is shown in 3(right).

Fig. 4(left) also shows that UDE model (1.2) achieves a very smooth fitting of the daily reported case data. More importantly, the fitting data is differentiable which implies that UDE model (1.2) can be regarded as advanced case data filtering or smoothing techniques with a great advantage over other soothing methods such as moving average, B-spline, expotential smoothing methods ([26]). The differentiable and fully observed fitting data can be directly used for subsequent data mining such as discovering first principle epidemic models to describe the transmission mechanisms, especially for data-driven techniques such as symbolic regression and sparse identification of nonlinear dynamical systems ([6]).

Moreover $\mathcal{R}_t$, represented as deep neural networks, can be directly integrated into other deep learning architectures and used for other purposes such as the effect of temperature, GDP, population density on infectious disease.

3.3. Comparison and Ensemble with EpiEstim and EpiNow2

In this part, we will compare deep learning method with 3 other methods state space, EpiEstim and EpiNow2 methods on the following aspects:

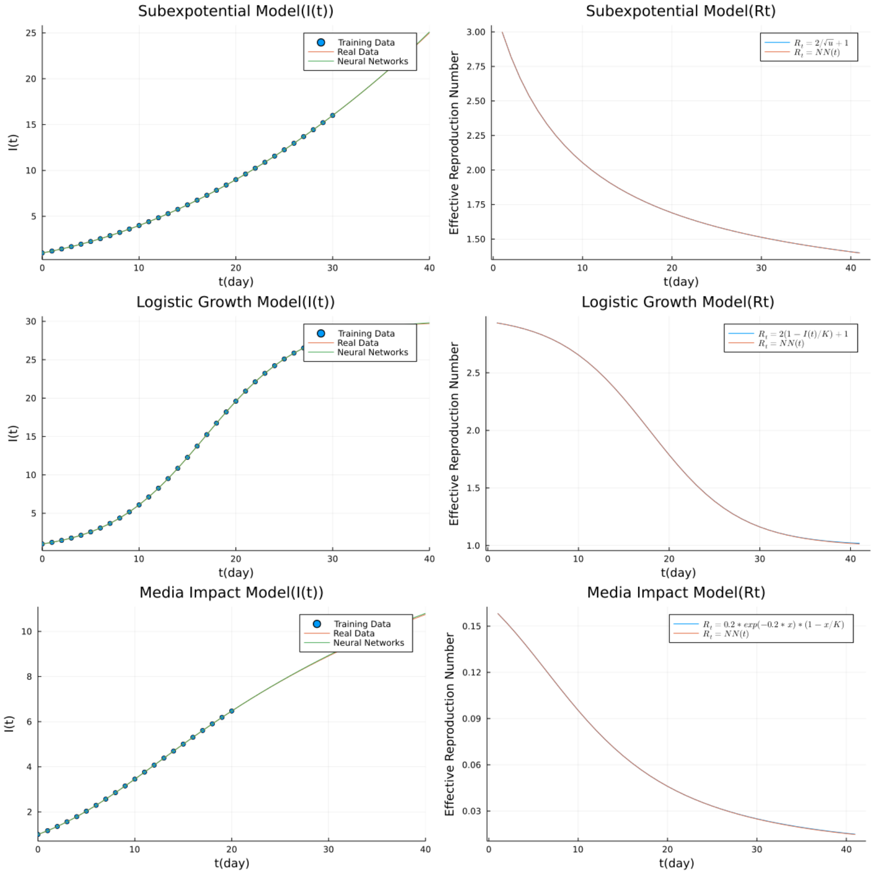

Figure 2. Left: fitting and predicting the toy models' generating data by universal differential equations model (1.2). Right: comparing real and trained effective reproduction numbers.

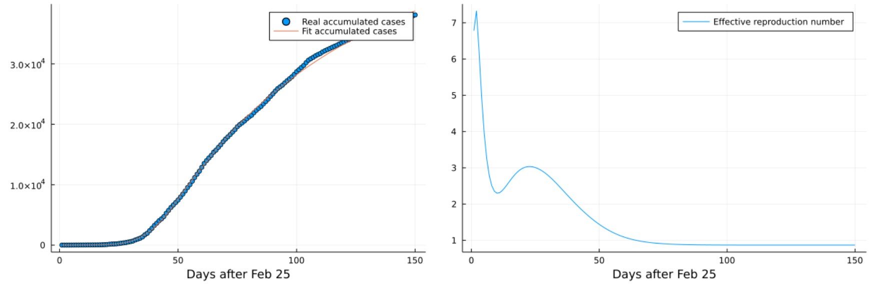

Figure 3. Left: Ontario's first wave COVID-19 case data fit by model (1.2). Right: effective reproduction number estimation by deep learning method.

- data sources: data needed to implement the method;
- smooth: describing the data fitting abilities, i.e., fitting the data but not over-fitting.
- speed: describing the algorithms speed.
- Accuracy of $\mathcal{R}_t$: the smooth properties and reliability (not bigger than 7) of $\mathcal{R}_t$,

and ensemble the four methods to estimate $\mathcal{R}_t$.

State-space method estimates effective reproduction number $\mathcal{R}_t$ based on the following state-space model:

$$\begin{cases} I(t) = (1-\gamma) * I(t-1) + \text{DailyCases}, \\ \dfrac{I(t) - I(t-1)}{I(t-1)} = \gamma(\mathcal{R}_t - 1). \end{cases} \tag{3.4}$$

Based on the daily confirmed case data and infection period data, kalman filter methods ([11]) can be used to estimate the $\mathcal{R}_t$ and handle the uncertainties. The ideas behind state space model (3.4) and UDE (1.2) are similar. They are both derived from classic SIR model ([27]). However, UDE model (1.2) embedded with neural networks can automatically incorporate many unknown features and export differentiable, fully observed fitting data and $\mathcal{R}_t$. UDE model (1.2) can be regarded as advanced state space model which is simple but incorporating many complex features.

EpiEstim method estimate $\mathcal{R}_t$ by the following equation ([12,41])

$$\mathcal{R}_t = \frac{\text{Incidence}_t}{\sum_{s=1}^{t} \text{Incidence}_{t-s} w_s},$$

where Incidence_t is the number of infection incidents on day t, and w_s is the serial interval. EpiEstim method needs daily confirmed cases data and serial interval data. Moreover, EpiEstim method can incorporate information on imported cases.

EpiNow2 method ([1]) is a sophisticated framework accounting for generation time distribution, incubation period distribution and delays in data such as reporting delays. Rather than reported onset date, EpiNow2 method estimates $\mathcal{R}_t$ on infection date, thus the effectiveness of public health intervention measures will be immediately reflected without delay. However, EpiNow2 method requires many high quality data sources and the algorithms are slow.

Fig. 4 shows the comparison and ensemble of these four methods with left and right parts describing the data fitting and $\mathcal{R}_t$ estimating results, respectively. It can be seen from Fig. 4(left) that both deep learning and EpiNow2 methods fit well but EpiNow2 method over-fits. Here, state space and EpiEstim methods do not fit daily confirmed cases data but directly calculate $\mathcal{R}_t$ from the case data. Fig. 4(right) shows that deep learning and EpiNow2 methods have good accuracy, while many $\mathcal{R}_t$ estimated by state space and EpiEstim methods are bigger than 7. Summary of comparisons is listed in Table 1.

Moreover, we integrate EpiNow2 and deep learning method in data fitting with equal weight (Fig. 4(left)), and integrate state space, EpiEstim, EpiNow2 and deep learning method in $\mathcal{R}_t$ with equal weight (Fig. 4(left)).

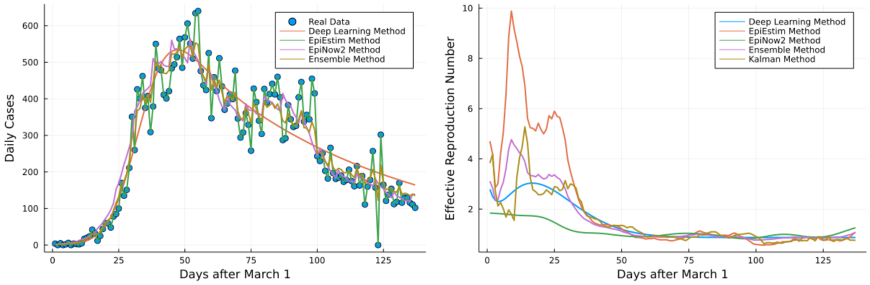

Figure 4. Left: Ontario's first wave COVID-19 case data fit by different methods. Right: effective reproduction number estimation by different methods.

Table 1. Comparison of different effective reproduction number estimation methods: deep learning, state space, EpiEstim, EpiNow2 Methods. Smooth measures the data fitting abilities, i.e., fitting the data but not over-fitting. Speed describe the algorithms speed. Accuracy of Rt measures the smooth properties and reliability (not bigger than 7) of $\mathcal{R}_t$. Tests are implemented on Ontario's first wave data and laptop Y7000P with i5-9300HF CPU, 16G RAM.

Methods	Data source	Smooth	Speed	Accuracy of $\mathcal{R}_t$
Deep Learning	Case data, infection period	strong	slow (3682s)	strong
State Space	Case data, infection period	weak	quick (1s)	weak
EpiEstim	Case data, serial interval	weak	quick (1s)	weak
EpiNow2	Case data, generation time, incubation period, delay distribution	normal	slow (2578s)	strong

4. Conclusion and Discussion

Effective reproduction number $\mathcal{R}_t$ is widely used in inferring the turning points and trend of an infectious disease, which is crucial for estimating the public health intervention measures and informing the public. In this paper, we use a deep neural network to represent $\mathcal{R}_t$ and embed it into a differential equation. This method is tested on three toy models, i.e., subexpotential, logistic growth and media impact models, and Ontario's first wave COVID-19 case data with good performance.

Compared with other $\mathcal{R}_t$ estimation methods such as compartment models ([19]), state space method, EpiEstim ([12]) method and EpiNow2 ([1]) method, deep learning method has the following advantages: (1) UDE model (1.2) is a simple model automatically incorporated with many complicated features and unknown mechanisms. Experts based compartment models, though without blackbox term and easy to interpret, are hard to capture unknown mechanisms of infectious disease, thus can not correctly estimate $\mathcal{R}_t$. Neural networks, though called as blackbox and hard to interpret, can incorporate complicated and unknown transmission mechanisms, thus can correctly estimate $\mathcal{R}_t$. (2) Deep learning method needs less data source. To accurately estimate $\mathcal{R}_t$, EpiNow2 method needs generation time, incubation period and reporting delays distributions data. Deep learning method only need infection period. (3) More importantly, neural network $\mathcal{R}_t$ can be directed integrated into deep learning ecosystem and used to discover more profound disease transmission mechanisms, such as the relationship between economy and infectious disease. Moreover, the obtained fitting data is differentiable and fully observed, and can be regarded as an alternative of daily confirmed cases with high quality, thus can be directly used for subsequent data mining such as unknown mechanisms discovering by symbolic regression and sparse identification of nonlinear dynamical systems ([6]).

Two obvious weakness of deep learning method are algorithms speed and uncertainties quantification. UDE method needs a lot of computation resources since one "backpropagation" (adjoint sensitivity analysis) means solving a big differential equation system. The speed weakness will become obvious especially when $\mathcal{R}_t$ is estimated in many regions and large time span. Some existing methods such as curriculum learning ([4]), transfer learning, decreasing testing data will increase the speed, but much more work are needed to improve the performances of UDE method. Handling uncertainty is another difficult job for universal differential equations with so many parameters. Some limited techniques exist such as bayesian neural differential equations ([14]), dropout ([21]), SGD as bayesian sampler ([32]). It will be our future work.

Acknowledgements

The authors would like to thank the referees for kind help to review the paper.

References

[1] S. Abbott, J. Hellewell, R. N. Thompson et al., *Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts*, Wellcome Open Research, 2020, 5(112), 112.

[2] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*, Cambridge University Press, 1991.

[3] A. G. Baydin, B. A. Pearlmutter, A. A. Radul and J. M. Siskind, *Automatic Differentiation in Machine Learning: A Survey*, J Mach Learn Res, 2018, 18.

[4] Y. Bengio, J. Louradour, R. Collobert and J. Weston, *Curriculum learning*, in *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*, ACM Press, Montreal, Quebec, Canada, 2009, 1–8.

[5] J. Bezanson, A. Edelman, S. Karpinski and V. B. Shah, *Julia: A fresh approach to numerical computing*, SIAM review, 2017, 59(1), 65–98.

[6] S. L. Brunton, J. L. Proctor and J. N. Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proc. Natl. Acad. Sci. U.S.A., 2016, 113(15), 3932–3937.

[7] Y. Cao, S. Li, L. Petzold and R. Serban, *Adjoint Sensitivity Analysis for Differential-Algebraic Equations: The Adjoint DAE System and Its Numerical Solution*, SIAM J. Sci. Comput., 2003, 24(3), 1076–1089.

[8] B. Chang, M. Chen, E. Haber and E. H. Chi, *AntisymmetricRNN: A Dynamical System View on Recurrent Neural Networks*, arXiv preprint arXiv:1902.09689, 2019.

[9] T. Q. Chen, Y. Rubanova, J. Bettencourt and D. K. Duvenaud, *Neural Ordinary Differential Equations*, in NeurIPS, 2018.

[10] X. Chen, *Ordinary differential equations for deep learning*, arXiv preprint arXiv:1911.00502, 2019.

[11] N. Chopin, O. Papaspiliopoulos et al., *An introduction to sequential Monte Carlo*, Springer, 2020.

[12] A. Cori, N. M. Ferguson, C. Fraser and S. Cauchemez, *A new framework and software to estimate time-varying reproduction numbers during epidemics*, Am. J. Epidemiol., 2013, 178(9), 1505–1512.

[13] M. Cranmer, A. Sanchez-Gonzalez, P. Battaglia et al., *Discovering Symbolic Models from Deep Learning with Inductive Biases*, ArXiv200611287 Astro-Ph Physicsphysics Stat, 2020.

[14] R. Dandekar, K. Chung, V. Dixit et al., *Bayesian Neural Ordinary Differential Equations*, ArXiv201207244 Cs, 2021.

[15] A. Davies, P. Veličković, L. Buesing et al., *Advancing mathematics by guiding human intuition with AI*, Nature, 2021, 600(7887), 70–74.

[16] E. De Brouwer, J. Simm, A. Arany and Y. Moreau, *GRU-ODE-Bayes: Continuous Modeling of Sporadically-Observed Time Series*, ArXiv190512374 Cs Stat, 2019.

[17] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, *Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding*, arXiv preprint arXiv:1810.04805, 2018.

[18] O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases*, John Wiley and Sons Ltd, Chichester, New York, 2000.

[19] O. Diekmann, J. A. P. Heesterbeek and M. G. Roberts, *The construction of next-generation matrices for compartmental epidemic models*, J. R. Soc. Interface, 2010, 7(47), 873–885.

[20] S. R. Dubey, S. K. Singh and B. B. Chaudhuri, *A Comprehensive Survey and Performance Analysis of Activation Functions in Deep Learning*, ArXiv210914545 Cs, 2021.

[21] Y. Gal and Z. Ghahramani, *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*, ArXiv150602142 Cs Stat, 2016.

[22] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT press, 2016.

[23] K. M. Gostic, L. McGough, E. B. Baskerville et al., *Practical Considerations for Measuring the Effective Reproductive Number, Rt*, PLoS Comput. Biol., 2020, 16(12), e1008409.

[24] K. He, X. Zhang, S. Ren and J. Sun, *Deep residual learning for image recognition*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 770–778.

[25] K. Hornik, M. Stinchcombe and H. White, *Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks*, Neural Networks, 1990, 3(5), 551–560.

[26] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, OTexts, 2018.

[27] W. O. Kermack and A. G. McKendrick, *A contribution to the mathematical theory of epidemics*, Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character, 1927, 115(772), 700–721.

[28] A. J. Kucharski, T. W. Russell, C. Diamond et al., *Early dynamics of transmission and control of COVID-19: A mathematical modelling study*, Lancet Infect. Dis., 2020, 20(5), 553–558.

[29] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning*, Nature, 2015, 521(7553), 436–444.

[30] Y. Lu, A. Zhong, Q. Li and B. Dong, *Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations*, in International Conference on Machine Learning, PMLR, 2018, 3276–3285.

[31] B. F. Maier and D. Brockmann, *Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China*, Science, 2020.

[32] C. Mingard, G. Valle-Pérez, J. Skalse and A. A. Louis, *Is SGD a bayesian sampler? well, almost*, Journal of Machine Learning Research, 2021, 22.

[33] M. Y. Niu, L. Horesh and I. Chuang, *Recurrent neural networks in the eye of differential equations*, ArXiv, 2019.

[34] K. V. Parag, *Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves*, PLoS Comput. Biol., 2021, 17(9), e1009347.

[35] A. Pinkus, *Approximation theory of the MLP model in neural networks*, Acta Numer., 1999, 8, 143–195.

[36] C. Rackauckas, Y. Ma, J. Martensen et al., *Universal Differential Equations for Scientific Machine Learning*, Preprint, In Review, 2020.

[37] Y. Rubanova, R. T. Q. Chen and D. Duvenaud, *Latent ODEs for irregularly sampled time series*, ArXiv, 2019.

[38] L. Ruthotto and E. Haber, *Deep Neural Networks Motivated by Partial Differential Equations*, J Math Imaging Vis, 2020, 62(3), 352–364.

[39] T. J. Sejnowski, *The unreasonable effectiveness of deep learning in artificial intelligence*, Proc. Natl. Acad. Sci. U.S.A., 2020, 117(48), 30033–30038.

[40] J. Sooknanan and N. Mays, *Harnessing social media in the modelling of pandemics-challenges and opportunities*, Bull. Math. Biol., 2021, 83(5), 57.

[41] R. N. Thompson, J. E. Stockwin, R. D. van Gaalen et al., *Improved inference of time-varying reproduction numbers during infectious disease outbreaks*, Epidemics, 2019, 29, 100356.

[42] J. Wallinga and P. Teunis, *Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures*, Am. J. Epidemiol., 2004, 160(6), 509–516.