SimNest: Social Media Nested Epidemic Simulation via Online Semi-supervised Deep Learning
(Supplementary materials)

Liang Zhao, Jiangzhuo Chen
Virginia Tech
liangz8@vt.edu, chenj@vbi.vt.edu

Feng Chen
University of Albany, SUNY
fchen5@albany.edu

Wei Wang, Chang-Tien Lu, Naren Ramakrishnan
Virginia Tech
iskatom@vt.edu, ctlu@vt.edu, naren@cs.vt.edu

Keywords-component; formatting; style; styling;

I. INTRODUCTION


In the following three sections, we first elaborate the partial derivatives of the loss function (in Equation 2 in the original paper) with respect to the weight matrix $W$. Then we describe the solution to the loss function $L_c$ in Section VI-B of the original paper. Finally, we introduce the settings of all the comparison methods in the experiment section of the original paper.

II. THE DERIVATIVES WITH RESPECT TO W

In this section, we elaborate the partial derivatives of the loss function (in Equation 2 in the original paper) with respect to the weight matrix $W$. This can be decomposed into the partial derivatives of each sub-loss function $L_1$, $L_2$, $L_3$, and $L_4$ in Equations 4, 6, 8, and 9, respectively.

$$\frac{\partial L_{3,u}}{\partial W_{j,k}} = (f_W(X_{u,t}) - Y_{u,t})s'(h^{(1)}_{j})W_{j}^{(2)}s'(h^{(2)}_{j})X_{i,k}$$

(1)

where $s'(x) = s(x) \cdot (1 - s(x))$.

$$\frac{\partial L_{1,u,t}}{\partial W_{j}^{(2)}} = (f_W(X_{u,t}) - Y_{u,t}) \cdot s'(h^{(1)}) \cdot s(h^{(2)})$$

(2)

where $L_{1,u,t} = L_1(f_W(X_{u,t}), Y_{u,t})$.

$$\frac{\partial L_{4,u}}{\partial W_{j,k}} = f_W(X_{u,t})s'(h^{(1)}_{j})W_{j}^{(2)}s'(h^{(2)}_{j})X_{u,t,k} - f_W(X_{u,t+1})s'(h^{(1)}_{j})W_{j}^{(2)}s'(h^{(2)}_{j})X_{u,t,k}$$

(3)

where $L_{4,u} = \sum_t T L_4(X_{u,t}, X_{u,t+1}, W)$.

$$\frac{\partial L_{2,t}}{\partial W_{j}^{(2)}} = \sum_u \left( \sum_v f_W(X_{v,t}) - \sum_v Q_v(p_E, p_I) \cdot s'(h^{(1)}_{u}) \cdot s(h^{(2)}_{u,j}) \right) \cdot X_{u,k}$$

(7)

Similarly, the derivative of $L_2$ is as follows:

$$\frac{\partial L_{2,t}}{\partial W_{j}^{(2)}} = \sum_u \sum_v f_W(X_{v,t}) - \sum_v Q_v(p_E, p_I) \cdot s'(h^{(1)}_{u}) \cdot s(h^{(2)}_{u,j}) \cdot X_{u,k}$$

(8)

III. THE SOLUTION TO THE LOSS $L_c$

In this section, we describe the solution to the loss function $L_c$ in Section VI-B of the original paper. Specifically, we solve the derivative with respect to $W$ and update the scaling parameter $\lambda_2$ alternatively.

The derivative of $L_3$ with respect to $W$ is deduced as follows:

$$\frac{\partial L_{3,u}}{\partial W_{j,k}} = \sum_t T \frac{\partial L_{3,u,t}}{\partial W_{j,k}}$$

$$\frac{\partial L_{3,u,t}}{\partial W_{j,k}} = \frac{\partial s(h^{(1)}_{j})}{\partial h^{(1)}_{j}} \frac{\partial h^{(1)}_{j}}{\partial h^{(2)}_{j}} \frac{\partial h^{(2)}_{j}}{\partial W_{j,k}}$$

$$= \sum_t T \left( \sum_i f_W(X_{u,i}) - p_I \right) s'(h^{(1)}_{i}) W_{j}^{(2)} s'(h^{(2)}_{i,j}) X_{u,t,k}$$

(5)

where $L_{3,u} = \sum_t T L_{3,u,t}$, and $L_{3,u,t} = L_3(X_{u,t}, W, p_I)$.

$$\frac{\partial L_{2,t}}{\partial W_{j}^{(2)}} = \sum_u \sum_v f_W(X_{v,t}) - \sum_v Q_v(p_E, p_I) \cdot s'(h^{(1)}_{u}) \cdot s(h^{(2)}_{u,j}) \cdot X_{u,k}$$

(7)

where $L_{2,t} = L_2(X_{t}, W, Z)$. 

$$\frac{\partial L_{2,t}}{\partial W_{j}^{(2)}} = \sum_u \sum_v f_W(X_{v,t}) s'(h^{(1)}_{u}) s(h^{(2)}_{u,j})$$

$$- \sum_u \sum_v Q_v(p_E, p_I) s'(h^{(1)}_{u}) s(h^{(2)}_{u,j})$$

(8)
The derivative of $L_c$ with respect to $W$ is as bellow:

$$\frac{\partial L_{c,i}}{\partial W_{j,k}(1)} = \sum_{l,t=a}^{L,a_i} \sum_{u} (\lambda_2 \alpha \cdot \sum_{l,p=a}^{L,a} \sum_{v} f_W(X_{v,p}) - C(i)) \cdot s'(h_t^{(1)}) W_j^{(2)} s'(h_{t,j}^{(2)}) X_{u,t,k}$$

where $\alpha = (a_e - a_s + 1)$.

$$\frac{\partial L_{c,i}}{\partial W_{j}^{(2)}} = \sum_{l,t=a}^{L,a_i} \sum_{u} \lambda_2 \alpha \sum_{l,p=a}^{L,a} \sum_{v} f_W(X_{v,p}) s'(h_t^{(1)}) s(h_{t,j}^{(2)})$$

$$- C(i) \sum_{l,t=a}^{L,a} \sum_{u} s'(h_t^{(1)}) s(h_{t,j}^{(2)})$$

In addition, the analytical solution of the scaling factor $\lambda_2$ is as follows:

$$\lambda_2 = \frac{\sum_i^{T'} M_i \cdot C(i)}{\sum_i^{T'} M_i^2}$$

where $M_i = (a_e - a_s + 1) \sum_{l,t=a}^{L,a_i} \sum_{u} f_W(X_{u,t})$.

### IV. SETTINGs OF COMPARISON METHODS

In this section, we introduce 6 competing methods. Among them, 4 methods are from social media mining: Linear Autoregressive Exogenous model (LinARX) [1], Logistic Autoregressive Exogenous model (LogARX) [2], Simple Linear Regression model (simpleLinReg) [7], Multi-variable linear regression model (multiLinReg) [6]. Another 2 methods are from computational epidemiology: SEIR [8] and EpiFast [3]. Their detailed settings are elaborated in the following.

1. **Linear Autoregressive Exogenous model (LinARX) [1], [9]**: This is a standard ARX model that builds the dependence of future visit percentage on the historical time series of CDC ILI visit percentage data [5] and volume of influenza tweets data $D_{(t)}$. The orders of LinARX for Twitter data time series and CDC time series are set as 2 and 3 respectively based on cross-validation. (2) **Logistic Autoregressive Exogenous model (LogARX) [2]**: On the basis of LinARX, this method add a logit function transformation on the historical time series to enforce the boundary 0-1 of the value of ILI visit percentage. The orders of LogARX for the two time series are both set as 2 based on cross-validation.

2. **Simple Linear Regression model (simpleLinReg) [7]**: This method assumes a linear mapping between the input, the volume of infectious tweets $D_{(t)}$, and the output, the future ILI visit percentages. (4) **Multi-variable linear regression model (multiLinReg) [6]**: This method treats a combination of keywords $K$’s volumes as a multivariate input of the simple regression model. (5) **SEIR [8]**: This model divides the population into four health states, namely susceptible (S), exposed (E), infectious (I), and recovered (R). The epidemic dynamics are modeled by ordinary differential equations. The visit percentage is calculated through multiplying the volume of the state “I” by a ratio, which is optimized by cross-validation. (6) **EpiFast [3]**: This model follows the definition in Section III, there are mainly two parameters that need to tune, $p_E$ and $p_I$. They are optimized by minimizing the error of the predicted and the actual ILI visit percentage via Neald Mead method [4].

### REFERENCES


