Team 4Sight Narrative:
Influenza Forecast for the 2015-2016 U.S. Flu season

Prithwish Chakraborty*† Jinghe Zhang*† Bryan Lewis‡
Sumiko R. Mekaru§ John S. Brownstein§¶
Madhav V. Marathe*‡ Naren Ramakrishnan*†

1 Model Narrative

For this year's Influenza forecasting challenge, Team 4Sight continues the use of dynamic data-driven SIRS model. This dynamic model is inspired from the works of Shaman et. al. [2] and aims to use a Bayesian Filter to continuously assimilate observed data sources into the model characteristics and generate an ensemble of models. A key distinguishing feature of our work will be the diversity of syndromic surveillance sources used. The spread of the ensemble predictions also reveals the underlying probability distribution of various seasonal characteristics such as start week and peak week. The model can be formally described as follows.

Let us denote the observed ILI percentage for the region of interest (including national level data) by y_t. We choose as a candidate the well defined SIRS model where S_t and I_t denote the number of people in 'Susceptible' and 'Infectious' compartment, at time t. Let us also denote the new infections moving into the I bucket at time t by $newI_t$ which can be directly computed from I_t. Let us denote the population size by N, the mean infectious period by D, the mean resistance period by L, and the basic reproductive rate at time t by $R_{0,t}$. Then the basic SIRS equation at time t can be given as

$$\begin{aligned} \frac{dS_t}{dt} &= \frac{N-S_t-I_t}{L} - \frac{\beta(t)I_tS_t}{N} - \alpha \\ \frac{dI_t}{dt} &= \frac{\beta(t)IS}{N} - \frac{I_t}{D} + \alpha \end{aligned} \tag{1}$$

where $\beta(t) = R_{0,t}/D$.

Let us denote a hidden layer of variables x_t that connect the SIRS model with the observed ILI percentages. The hidden variable set can be thought of

*Dept. of Computer Science, Virginia Tech, USA
†Discovery Analytics Center (http://dac.cs.vt.edu), Virginia Tech, USA
‡Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Tech, USA
§Childrens Hospital Informatics Program, Boston Childrens Hospital, USA
¶Department of Pediatrics, Harvard Medical School, USA

as an n-tuple x_t, as

$$x_t = (S_t, I_t, R_0, D, L, f, r)$$

The equations governing the Bayesian filter can be given as:

$$\begin{array}{rcl} y_t & = & f * newI_t + \mathcal{N}(0, r) \\ x_t & = & g(x_t | x_{t-1}) \end{array} \tag{2}$$

where g denotes the dynamic model transition from time $t-1$ to t.

g can be a general purpose transition function. For our purpose, we perturb S and I via the SIRS equation and the remaining state parameters using a random walk model within specified bounds. For this first round of submissions, we used a particle filter with 10000 particles to fit the filter. The particle filter involves systematic restarts and re-orientations to introduce jitter and handle particle degeneracy. The distribution of the particles provide the posterior distribution over the SIRS parameters and can be used to directly infer the parameters.

2 Using surrogate Sources

The method so described above can be thought of as a general purpose algorithm where we can introduce information about different sources by modifying Equation 2. For this first submission, we only used Google Flu Trends data. Additional data sources such as weather, Google Search Trends, Healthmap, Wikipedia, and atmospheric ifnromation will be utilized for future submissions. This surrogate information was used to modify the transition equation for other latent variables such as R_0 as:

$$R_{0,t} = R_{0,t-1} + \mathcal{N}(0, cov(GFT_{t-1}, GFT_t)) \tag{3}$$

Following Chakraborty et al. [1] we intend to analyze a myriad of data sources to train a more precise model with lower uncertainty bounds.

References

[1] Prithwish Chakraborty, Pejman Khadivi, Bryan Lewis, Aravindan Mahendiran, Jiangzhuo Chen, Patrick Butler, Elaine O. Nsoesie, Sumiko R. Mekaru, John S. Brownstein, Madhav V. Marathe, and Naren Ramakrishnan. Forecasting a moving target: Ensemble models for ILI case count predictions. In *Proceedings of the 2014 SIAM International Conference on Data Mining, Philadelphia, Pennsylvania, USA, April 24-26, 2014*, pages 262–270, 2014.

[2] Jeffrey Shaman, Virginia E Pitzer, Cécile Viboud, Bryan T Grenfell, and Marc Lipsitch. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS biology*, 8(2):e1000316, 2010.