Stochastic Markov Disease Modelling

Learn Imagine Adapt decided to apply advanced modelling methods to predict a suitable infection rate for our app users. Using historical past data, coupled with real time updated data from our users, our embedded model can predict a certain infection rate for each user.

The **Stochastic Markov Disease Model** or SMDM is designed so that when information is fed in, the model can easily update its parameters to account for random noise or a sudden surge in disease spread. Our model's algorithms are developed by blending traditional **Markov Chain Modeling**, **Bayesian Modelling** and **Reinforcement Learning** algorithms.

To make the model more cohesive, we have split it into 3 parts:

1. Predicting the infection rate from real-time and historical data
2. Predicting a person's travel patterns from historical data
3. Predicting how an infectious disease will spread using real-time data.

By splitting it up into 3 sections, the model can easily learn its parameters and become more accurate, hence making it a reinforcement learning model.

Mathematical Underpinnings

The model's mathematical algorithm consists of a blend of the 3 algorithms as mentioned before.

1. **The prediction of the infection rate:**

$$f(A,\mu_A,\sigma_A) = \frac{1}{\sigma_A\sqrt{2\pi}} e^{-\frac{1}{2\sigma_A^2}(A-\mu_A)^2}$$

$$C = g(A,\mu_A,\sigma_A,\lambda_m) = \lambda_m \frac{f(A,\mu_A,\sigma_A)}{f(\mu_A,\mu_A,\sigma_A)}$$

$$L = 1 - l = C_L$$

$$h(C,n) = C^T J_n(J_n - I_n)$$

$$q(C,n) = C I_n + \frac{(1-C)h(C,n)}{\sum_{row} h(C,n)}$$

$$P_t = \prod_{k=1}^{d} q(C_k,n)$$

$$MD_t = [MD_{t-1} + (\lambda_{mb} MB_t - MD_t)] \times P_t$$

$$AR_t = CDF\left(\frac{MD_t}{Sus_t}\right)$$

$$\boxed{[S \rightarrow I]_t(\lambda_{ir}) = \lambda_{ir}\, I_n AR_t \frac{e^{\frac{Inf_t}{Population}} - 1}{e - 1}}$$

2. **The prediction of people's travel patterns:**

$$TR = Road_{ratio} + Water_{ratio} + Air_{ratio}\ (1.6, 0.4, 0.1)$$

$$ga(G,M_t,\lambda_g) = \lambda_g \frac{M_t}{\mathcal{E} + \sum_{row} G}$$

$$R_t = mg(G,M_t,\lambda_g) = M_t + ga(G,M_t,\lambda_g) - \frac{G\lfloor[1-G]\rfloor}{\sum_{row}\lfloor[1-G]\rfloor}\sum_{row} ga(G,M_t,\lambda_g)$$

$$\boxed{[S \rightarrow S]_t = (I_n - [S \rightarrow I]_t)\, R_t}$$

3. <u>**The prediction of infectious disease spread:**</u>

$$S_t = [S_{t-1} + (\lambda_b B_t - D_t)] \times P_t$$

$$MSIR_t = \begin{bmatrix} [S \to S]_t & [S \to I]_t & 0 \\ 0 & [I \to I]_t & [I \to R]_t \\ 0 & 0 & [R \to R]_t \end{bmatrix}$$

$$SIR_t = [SIR_{t-1} + (\lambda_b SIRB_t - SIRD_t)] \times MSIR_t$$

Parameter Bayesian Updating

As you can see from the mathematical algorithms, there exists some parameters that are allowed in **green**. These are the parameters that were learnt from historical and real-time data.

After applying machine learning models using **DataRobot** and other methods, we discovered the following interesting trends:

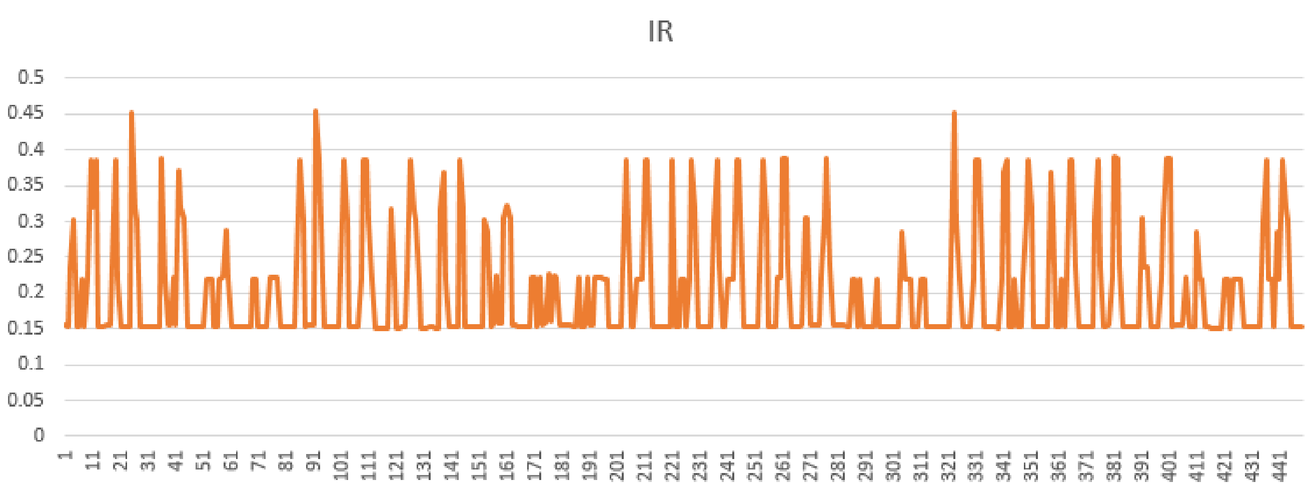
Variable	**Data Source (data transformed)**	**Model Parameter Importance**	**Rationale**
Train Congestion Rate	Open Transport Data - OPAL	**+15.183**	We found that as train congestion increased, the infection rate suddenly increased. This is clearly due to the fact that people are closer together in trains, and thus spreading of diseases is most optimal there.

Bus Congestion Rate	Open Transport Data – Bus Network	**+10.233**	Likewise, fascinatingly enough, as bus congestion increased, so did the infection rate. This mirrors the train congestion trend, as clearly, as bus passengers are increased, more diseases are easily spread.
Population Proximity Movement Rate	ABS and ATO data HERE Maps train, road connections data	**+9.123**	We noticed as the connections to one city or suburb increases, so does the infection rate. We also noticed that larger population areas tend to have more movement of people, and thus it can cause an increase again in infection rates.
Time of Day	ALL DATA	**+8.963**	We noticed that at the peak hours, the infection rate is booming. This is so, as the time of day is directly correlated and related to train and bus congestion, and thus it is reasonable to conclude that peak hours harbor more disease spreading potential.
Air Quality Index	NSW Air Quality Index data	**+7.343**	We found as the air quality degraded, the infection rate increased by an amount. We noticed that this was because people become weaker as the air quality is a direct causer of poor health. As the AQI increases (or AQ worsens), the infection rate increases.

Standardized Capital and Total Income	ABS Income data	**+5.234**	We also found that as your income increases, you tend to develop a lower chance of getting a disease. This is so, as we noticed that this was correlated with the health status of people. HOWEVER, we also found that very high-income people might develop a higher change of getting a disease. We noticed that these people tend to travel in public transport more often, and this could be a reason for this trend.

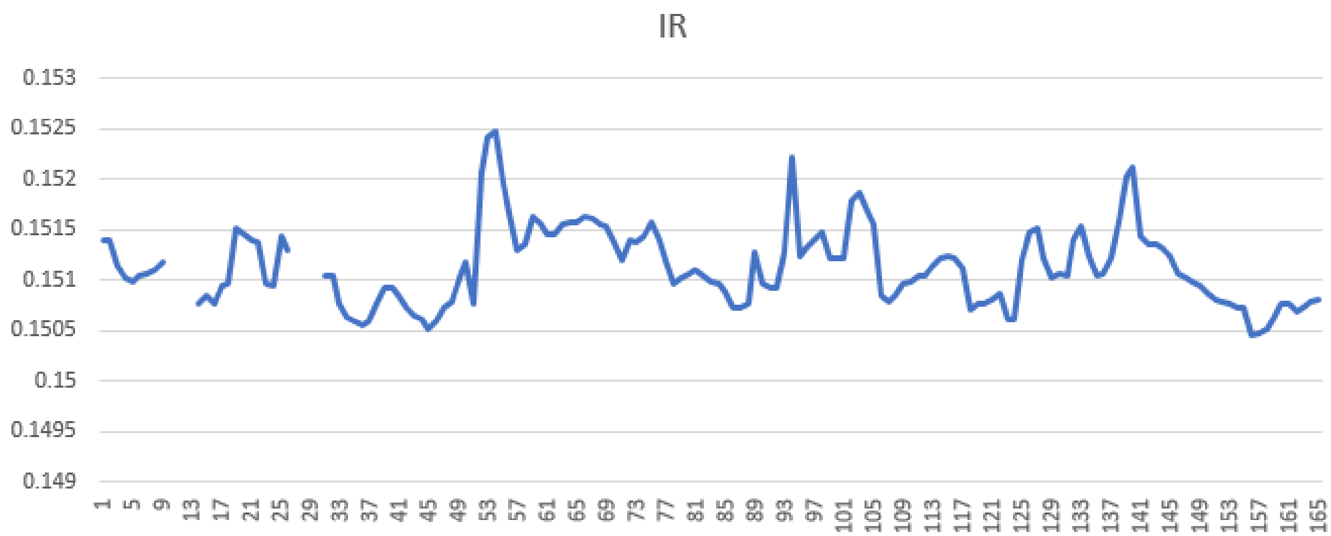
Model results

Let us inspect Blacktown's infection rate over the month of February 2017:



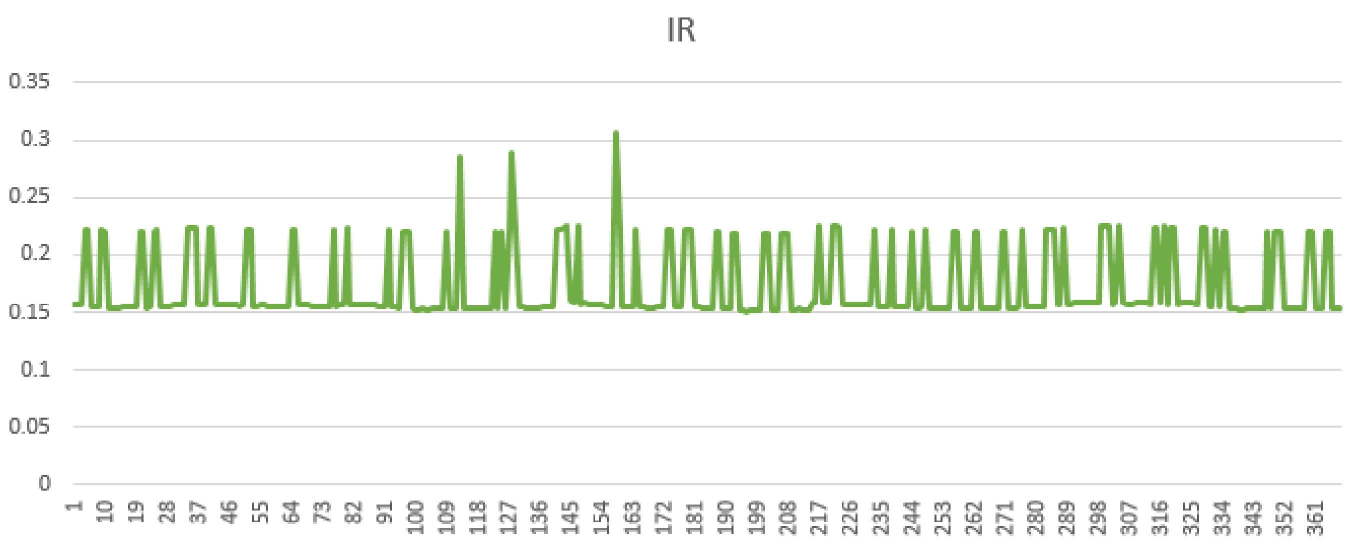
Clearly, you can see a stochastic trend. You can see the dips and increases which symbolize the onset of nights and day cycles. A sharp increase is seen in the midday.

Richmond's infection rate pattern is different:



This was since Richmond is further away from the city, and so spreading to there is lower (hence average rate of only 0.15%)

To confirm our suspicions, let's check Liverpool:



As you can see, Liverpool's average is around 0.225, but a more cyclical trend is seen.

Our team Learn Imagine Adapt thanks you!