

Benchmark Data Generation from Discrete Event Contact Network Models

Emmet Beeker¹, M.S., David W. Bauer¹, Ph.D., Mojdeh Mohtashemi^{1,2}, Ph.D.

¹The MITRE Corporation; ²MIT CS and AI Lab

OBJECTIVE

The objective of this study is to generate benchmark data from a discrete event model simulating the transmission dynamics of an infectious disease within and between contact networks in urban settings using real population data. Such data can be used to test the performance of various temporal and spatio-temporal detection algorithms when real data are scarce or cannot be shared.

BACKGROUND

Historical data are essential for development of detection algorithms. Spatio-temporal data, however, are difficult to come by due to variety of issues concerning patient confidentiality. Several approaches have been used to generate benchmark data using statistical methods [1]. Here, we demonstrate how to generate benchmark data using a discrete event model simulating inter- and intra-contact network transmission dynamics of infectious diseases in space and time using publicly available population data.

METHODS

An agent-based, stochastic SEIR [2] model of transmission dynamics of an infectious disease is developed using real population data [3]. The simulation environment is partitioned into contact networks such as households, work places, and schools each with its own disease parameters. Individuals (agents) can be members of more than one network. In a network, an individual belongs to an epidemiologic stage of a disease, such as susceptible, exposed, infectious, and removed for a period dictated by disease specific parameters after which may move to another. The rates at which members move from one stage of disease to another are governed by disease-specific epidemiologic parameters including the infectivity, latent, and recovery periods, average number of contacts, and effective transmission rate. The disease is propagated through the population by means of infectious individuals moving between contact networks. An important epidemiologic parameter in the model is the basic reproductive number R_0 , which is the average number of secondary cases generated by an infectious case in an entirely susceptible population. Thus R_0 signifies the disease potential to become epidemic. Mathematically, $R_0 = \beta DS_0$, where β is the effective transmission rate, D is the average infectivity period, and S_0 is the size of the susceptible population when the infection is introduced into the population. The model captures human behavior, where individuals can choose dynamically to stay home, go to work, seek medical attention, or which medical facility to go to when symptomatic. The output of the model consists of spatio-temporal data corresponding to the time of visit and patient location (census tracts) when symptomatic patients show up at a medical facility.

RESULTS

The time series data generated here is based on population information for the city of Chicago, which was used to locate homes, work places and schools in the scenario

based on census tract information. Disease data are then collected at medical facilities located within the scenario area. Figure 1 illustrates the simulated data mimicking the influenza season for $R_0=1.4$ (a plausible value for influenza epidemics), $\beta = 4 \times 10^{-5}$, $D=3.5$ days, and population size of 10,000 for the catchment area of one medical facility in Chicago. Here stochasticity is introduced by modeling the probability of an agent becoming ill as a function of the size of the contact group. In the first time series basic human behavior is captured in which agents stay home when symptomatic, while the second time series in addition captures the behavior of *worried well* (here 0.5% of the population plotted in green). These are agents with symptoms similar to the disease who are infected with another disease. While they introduce noise in the data, they act to mitigate the spread of the disease through social distancing. The percentage of worried well individuals tends to increase during epidemic or pandemic situations.

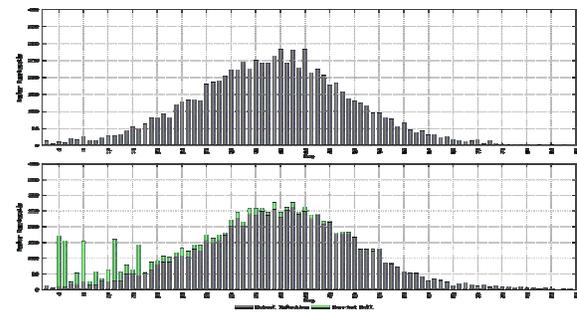


Figure 1: Synthetic influenza-like epidemic trajectory generated by the model.

CONCLUSIONS

We proposed a new technique for generating benchmark data for testing the performance of various detection techniques. The value of this model over those created under statistical assumptions is three-fold: *i) infection transmission dynamics in time, space and across contact networks is modeled using real population data, ii) the health seeking behavior of a population is accounted for, and iii) different diseases can be simulated within the same modeling framework.* For near future, we plan to make parameter variability explicit and more systematic across contact networks, generate spatial data, apply space-time detection techniques to the data, and compare their performance measures. We will make this data generation software available for public use.

REFERENCES

- [1] Kulldorff M., et al. Benchmark data and power calculations for evaluating disease outbreak detection methods. *MMWR Morbidity and Mortality Weekly Report* 2004, 53:144-151.
- [2] Ferguson N, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 2005, 437:209-214.
- [3] U.S. Census Bureau. United States Census 2000. <http://www.census.gov/main/www/cen2000.html>.