Epidemic Alert & Response Framework and Technology Based on Spreading Dynamics Simulation

Chunhua Tian\textsuperscript{1}, Xun Zhang\textsuperscript{2}, Wei Ding\textsuperscript{1}, and Rongzeng Cao\textsuperscript{1}

\textsuperscript{1} IBM China Research Laboratory, Beijing, China
\{chtian,dingw,caorongz\}@cn.ibm.com
\textsuperscript{2} China Academy of Science, Beijing, China
zhangx04@mails.gucas.ac.cn

Abstract. There have been many research works on epidemic spreading by mathematical or simulation methods. However, to make it practical as a decision support tool for policy maker, it should be combined with business operation of epidemic control. Thus, an alert and response framework covering outbreak detection, policy analysis and policy execution is proposed in this paper. Key technologies for such a framework at modeling, analysis and data collection aspects are discussed. Sensitivity analysis is proposed for model verification and confidence analysis. Critical point identification is another important topic for control policy making by network topology analysis. The importance and feasibility of these technologies are illustrated by the experiment on Beijing SARS (Severe Acute Respiratory Syndrome) data in 2003.

Keywords: Epidemics Simulation, Epidemic Alert & Response, Sensitivity Analysis, Critical Point Identification.

1 Introduction

Effective spreading control is an important approach in fighting with epidemics. At present, epidemics spreading control policy making is still heavily based on heuristic experience in past epidemic control. There is a big gap to be an effective control. On one hand, past experience is not always applicable to current cases due to diseases evolvement, social behavior change, transportation network improvement, and other social and environmental changes. On the other hand, effective epidemic control needs more precise estimations besides heuristic experience. So there is a great need of quantitative methods for policy maker in spreading control, such as spreading pattern analysis, critical control point through spreading and social network analysis, and policy comparison through what-if or scenario analysis. However, current research mostly focus on spreading process prediction by mathematical or simulation methods, which is not closely aligned with business operation of epidemic control. The purpose of this paper is to analysis the technology requirement and gap for the business operation and propose a technology framework for future work.

2 Literature Review

Bagni et al. [1] give an exhaustive review about the models and recent work in epidemics study. So only a brief review from model structure and research approach perspective is
Epidemics models can be distinguished from different perspectives: 1) continuous model where the total population can be partitioned infinitely, or discrete model with individual as the smallest grain, 2) pure temporary model, or spatial-temporary model, and 3) stochastic/determined model. SIR (Susceptible, Infective, and Recovered) model is the start point of many other continuous models for epidemic spreading, such as SIS, SEIR (‘M’ denotes ‘Recovered’), MSEIR (‘M’ denotes “maternal antibody protection”), etc. All these models assume that the population is "well mixed" (i.e., either not distributed or geographically distributed in a uniform manner). With the application of computer simulation, more complicated models are proposed considering spatial dimension, traffic network, diversity of population and many other important factors. For example, global epidemics spreading model through air travel infrastructure is proposed by Colizza et al. [2].

As the complexity of the epidemic process, simulation becomes the trend in epidemiology study compared with traditional analytical approach. Epidemic simulation software is closely related with epidemic models. The first generation epidemic software (such as WinEpi, Vensim) provides fruitful mathematical models support. The second generation epidemic software enhances mapping and spatial analysis functions. The typical software include Epimap, HealthMapper, SIGEpi, etc. The third generation epidemic software has more considerations to the impact of social and environmental factors. EpiSims [3] is an agent-based modeling platform based on individuals everyday behaviors. A disadvantage of this approach is that it requires large amounts of computing resources even for small models. To avoid such disadvantage, STEM is developed by Ford et al. [4], whence continuous modeling is used for large areas while agent based model is applied in more specific areas.

3 Epidemic Alert & Response Framework and Technology

3.1 Alert & Response Framework

Most epidemic research just focus on epidemic spreading process prediction, without clear connection with business operation of epidemic spreading control. What policy maker really needs is a decision supporting tool in the whole lifecycle. Analytical mechanism is needed for early alert about epidemic outbreak. When there is an epidemics outbreak, policy makers often want to know what critical point for epidemics control is, what the effectiveness of different control policies is, what will happen if the real status is different from the reported scenarios, and etc. Besides policy making, the execution of policy is also very important. For example, how to effectively allocate healthcare workers and drugs to disaster areas.

Based on the business operations of epidemics spreading control, an Alert and Response framework shown in Fig. 1 is proposed which covers all three stages of epidemics control. Monitoring system acts a fundamental infrastructure for the analysis modules and GIS (Geographical Information System) display. Buckeridge et al. [5] gives a detailed summary about how to use spatial and other covariate information from disparate sources to improve the timeliness of outbreak detection. In operation research there have been lots of works on execution planner. So this paper focuses on the other 2 blocks: Policy/Scenario Analyzer and Epidemics Dynamics Simulation.
The technologies framework can be arranged into 4 layers: data, model, application, and representation layer, as illustrated in Fig. 2.

3.2 Data & Model Layer

Data is the fundamental layer to the whole framework. Unfortunately not all data are available from public channels or epidemics monitoring information system. Interview or other novel techniques for data collection/generation may be needed. For example, in EpiSims, the travel behaviors are generated by TRANSIS, which in turn relies on interview. Interview usually requires onerous effort. Some automatic approaches may be developed based on novel technologies. For example, cell-phone positioning technology is potential to be adopted in generating travel behavior or traffic flow data.

For modeling, a big challenge is how to obtain a precise estimation of model parameters. In epidemics spreading process, new epidemics data are continuously available with the development of disease. So the adaptive parameter estimation method is more desired. Furthermore, even if the monitoring system is full-scale, not all the states are observable, such as people exposed or in incubation period. State observer technology from control theory may be used to estimate the parameters or hidden states.
Another big challenge for modeling is social network construction, which is closely related with epidemics transmission mechanism. For example, some kinds of diseases (e.g., SARS) are diffused within one type of host (e.g., through human contact), traffic flow data may be enough to construct social network. While for the diseases (e.g., H5N1 Avian Influenza Virus) that diffused across hosts (e.g., through animal and human contact, rather than direct transmission through people), it is much harder to construct the social network.

3.3 Application Layer

3.3.1 Key Analysis Functions

Based on the model and data, various applications can be developed for decision support, such as epidemic spreading prediction, critical point identification, and policy/scenario analysis.

Epidemic spread prediction means predicting the trend and pattern of epidemic spread in the spatial and temporal dimension. However it is hard to have precise forecasting about number of each class, since quantitative model is an approximation to reality at the best, and the parameters are also not the exact number. For decision support purpose, spreading comparison under different scenarios and policies are more important instead.

Contact network is a key factor in epidemics spreading. There have been some researches on epidemiological process over complex network. Network property analysis (e.g., average clustering coefficient, characteristic path length) can provide some guidance (e.g., [3]) in critical point/approach identification for epidemic surveillance. Besides network analysis, rule-based or pattern matching technology may also be useful to help policy maker identify critical path in the large data source.

3.3.2 Sensitivity Analysis

Epidemic simulation often faces great challenge from policy makers about the confidence of simulation result, since epidemics model is an approximation to the real process, and parameters are just estimated value. There are fundamental technical difficulties in such model verification in philosophy just like most social studies. Control experiment is impossible, and spreading process is not replicable. So sensitivity analysis is proposed in this paper to analysis confidence level of the simulation result.

Sensitivity analysis investigates how model output changes as the result of the change of input parameters. Uncertainty in parameter estimation can be quantified under different assumptions (e.g., heterogeneity in variance, correlated errors, etc.). Sampling techniques (e.g., Latin hypercube, simple random sampling) are useful to explore the parameter space and assess the uncertainty of epidemiological quantities.

The real spreading process $X^r$ over geography $p$ and time $t$ depends on initial state $X^r_0$ (i.e., scenario). It can be expressed as the model with “optimal” parameter vector $\alpha^r$ with a un-modeled process $\varepsilon$, shown as

$$X^r(p, t) = f(p, t; X^r_0, \alpha^r) + \varepsilon(p, t).$$  

(1)
\( f(p, t; X^r_0, \alpha^r) \) is the best approximation of the spreading process according the given model. Before all the data are available, it is usually hard to obtain. A guess parameter to \( \alpha^r \), \( \alpha^e \) and guess to initial states \( X^r_0 \), \( X^e_0 \) are obtained according to partial epidemics data. Hence the result for the simulation model is expressed as

\[
X^e(p, t) = f(p, t; X^e_0, \alpha^e).
\]  

(2)

Although it is impossible to obtain the real initial state and optimal parameter, it is still useful to know the sensitivity of spreading process \( X(p, t) \) on these parameters. If the sensitivity is low, the simulation results are still believable.

Denote \( \beta \) as the joint vector of \( X_0 \) and \( \alpha \), i.e., \( \beta = [X_0; \alpha] \). For \( X_0 \in \delta(X^e_0) \) and \( a \in \delta(a^r) \), calculate spreading process \( X(p, t) = f(p, t; X_0, \alpha) \) by simulation. The confidence level of simulation result can be defined as

\[
\lambda = 1 - \eta/(1 + \eta).
\]  

(3)

Whence, the variance of spreading process, \( \eta \), is defined as

\[
\eta = \max_{\beta \in \delta(\beta)} ||X - X^e|| / X^e.
\]  

(4)

\( \| \) is a measure of function vector. Operator ‘/’ means the elementary division, i.e., for \( a, b \in R^n \), \( a / b = [a_1 / b_1, \ldots, a_i / b_i, \ldots, a_n / b_n]^T \in R^n \).

The sensitivity of a parameter \( \beta_i \) in vector \( \beta \) around \( \beta^r \) can be defined as

\[
S(\beta_i)|_{\beta^r} = \frac{\partial ||X - X^e||}{\partial \beta_i},
\]  

(5)

if such a partial differential exist. If it dose not exist, the sensitivity can still be evaluated by sampling difference calculation.

4 Experiment

In this section, SARS spreading in Beijing, 2003 from Chinese Minister of Health [6] is adopted to illustrate the importance of some key research topics in Section 3.

4.1 Model and Prototype

The BloComp(2,7) model proposed by Zhang et al. [7] is adopted here. The model has 2 types of environment according to contact rate: free environment and isolated environment. Besides the 4 states in SEIR model, 3 more states are extended, including “P” as Possible infectious, “D” as Diagnosed infectious and “H” as Health worker. A prototype based on Repast [8] is developed.
4.2 Parameter Estimation

The parameters and functions are given as the following formulas as [7],\( \varepsilon = 0.04 \), \( d_{ep} = 0.16 \), \( d_{iq} = 1/3 \), \( d_{sp} = 0.027478 \), \( b_{sp} = 0.06341 \), \( d_{pq} = 0.12197 \), \( \gamma = 1/30 \), \( \alpha = \alpha_1 = 0.14/30 \), \( f(S,E,I,R) = \lambda_j(t)(0.1E + I) \), and \( g(H,P,D) = \lambda_g(t)D \).

However, in estimating \( \lambda_j(t) \) and \( \lambda_g(t) \), our approach differs a little from the approach of Zhang et al. [7], where they assume that today’s number of new diagnosed equals to 8-day-ago’s num of infective.

Let \( \Delta d(t) \) denote the number of new diagnosed minus the number of new infected health care workers. Under the assumption that the average transition time from exposed class to the infective class is 5 days and that from infective class to the D-class is 3 days, the exposed at time \( t \) is the \( \Delta d \) from \( t-5 \) to \( t-1 \), and the number of infectious at time \( t \) is the \( \Delta d \) from t-8 to t-6. Thus \( \lambda_j(t) \) can be estimated as

\[
\lambda_j(t) = \frac{\Delta d(t)}{0.1 \sum_{j=1}^{5} \Delta d(t-j) + \sum_{j=6}^{8} \Delta d(t-j)} (t \geq 8) .
\]

(6)

\( \lambda_g(t) \) can be also obtained in a similar way. Here a curve fitting method is adopted since it has more extensibility to other virus. Upon the observation of 16 days’ data from Apr. 25th to May. 10th, Lorenz curve is selected which is given as

\[
y(t) = y_0 + \frac{2A}{\pi} \cdot \frac{w}{(t - t_c)^2 + w^2} ,
\]

(7)

Whence, \( y_0 \), \( w \), \( A \) and \( t_c \) are parameters to be estimated.

![Fig. 3. Parameter Estimation with 10, 16, 20, and 50 Days’ Historical Data](image)
To study the impact of available data on the prediction precision, we use 10, 16, 20 and 50 days’ data respectively to estimate the incident rate function $\hat{\lambda}(t)$. The results are shown in Fig. 3. The estimation changes apparently with available data, which means that adaptive parameter estimation is usually necessary for real prediction.

Then, based on the estimation from 16 days’ data, the simulation result about infectious healthcare worker and diagnosed class are shown in Fig. 4. These figures show that the simulation results have the same trend with the real data, however with significant error, especially around the jumping point such as April. 29 shown in Fig. 4(a). However these kinds of jumping points may be due to the imperfection of data reporting mechanism, which is also one key challenge for epidemics prediction.

![Fig. 4. Simulation Result (solid curve) vs. Data from Chinese Minister of Health (dotted curve)](image)

(a) Num of Incidence of Healthcare Workers  (b) Num of Diagnosed Class

### 4.3 What-If Scenario Analysis

Fig. 5 shows the spreading process under three different scenarios. Scenario 1 is the simulated process based to the estimation from real data. Scenario 2 is the spreading process under lightly looser control measures with 3 times incidence of the original simulation. Scenario 3 is much looser with 10 times incidence of the original simulation. It clearly shows that quarantine policy does matter to the spreading process.

![Fig. 5. Spreading Process under Three Control Policies](image)
5 Conclusions

In a world reeling from the resurgence of old diseases and the emergence of new diseases, epidemic alert and response is very important. Current epidemics study need move from prediction to decision support. This paper proposes a research framework to identify the key analytical methods to fill such gaps. The experiment on 2003 Beijing SARS data proves the importance of those technologies. There are many interesting topics for future exploration, such as robust and adaptive parameter estimation method, critical control path identification algorithm, new data collection method, flexible simulation platform for general epidemics study, etc.

References

1. Bagni, R., Berchi, R., Cariello, P.: A comparison of simulation models applied to epidemics. Journal of Artificial Societies and Social Simulation. 3 (2002).
2. Colizza, V., Barrat, A., Barthélemy, M., Vespignani, A.: Prediction and predictability of global epidemics: the role of the airline transportation network. Proceedings of the National Academy of Science of the United States of America. 7 (2006) 2015-2020.
3. Eubank, S., Guclu, H., Anil Kumar, V. S., et al.: Modelling disease outbreaks in realistic urban social networks. Letters to Nature 429 (2004) 180-184.
4. Ford, D. A., Kaufman, J. H., Eiron, I: An extensible spatial and temporal epidemiological modeling system. International Journal of Health Geographics. 4 (2006).
5. Buckeridge, D. L., Burkom, H., Campbell, M., et al.: Algorithms for rapid outbreak detection: a research synthesis. Journal of Biomedical Informatics. 38 (2005) 99-113.
6. Chinese Minister of Health. <http://168.160.224.167>.
7. Zhang, J., Lou, J., Ma, Z., Wu, J. H.: A compartmental model for the analysis of SARS transmission patterns and outbreak control measures in China. Applied Mathematics and Computation. 162(2005) 909–924.
8. Repast Organization for Architecture and Development. (2003). http://repast. sourceforge.net.