Expert-in-the-Loop Prescriptive Analytics using Mobility Intervention for Epidemics

Joon-Seok Kim
George Mason University
jkim258@gmu.edu

Hyunjee Jin
George Mason University
hjin6@gmu.edu

Andreas Züfle
George Mason University
azufle@gmu.edu

ABSTRACT

Due to complexity of social phenomena, it is a big challenge to predict the curves of epidemics that spread via social contacts and to control such epidemics. Misguided policies to mitigate epidemics may result in catastrophic consequences such as financial crisis, massive unemployment, and the surge of the number of critically ill patients exceeding the capacity of hospitals. In particular, under/overestimation of efficacy of interventions can mislead policymakers about perception of evolving situations. To avoid such pitfalls, we propose Expert-in-the-Loop (EITL) prescriptive analytics using mobility intervention for epidemics. Rather than employing a purely data-driven approach, the key advantage of our approach is to leverage experts' best knowledge in estimating disease spreading and the efficacy of interventions which allows us to efficiently narrow down factors and the scope of combinatorial possible worlds. We introduce our experience to develop Expert-in-the-Loop simulations during the *Challenge on Mobility Intervention for Epidemics*. We demonstrate that misconceptions about the causality can be corrected in the iterations of consulting with experts, developing simulations, and experimentation.

CCS CONCEPTS

• **Information systems → Geographic information systems**;
• **Computing methodologies → Agent / discrete models**.

KEYWORDS

expert-in-the-loop, prescriptive analytics, epidemic modeling, microsimulation, mobility, intervention

ACM Reference Format:
Joon-Seok Kim, Hyunjee Jin, and Andreas Züfle. 2020. Expert-in-the-Loop Prescriptive Analytics using Mobility Intervention for Epidemics . In *Prescriptive Analytics for the Physical World (PAPW '20), August 24, 2020, San Diego, CA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3347146.3363350

1 INTRODUCTION

Predicting human behaviors and the resulting spread of a pandemic is a tremendous challenge due to complexity (and often irrationality) of humans. Physicist Murray Gell-Mann, Nobel laureate who conceived the quark, famously said "Imagine how hard physics would be if electrons could think" [3]. To predict pandemic spread the atoms that comprise the model for simulation and predict are indeed humans. The difficulty of predicting disease pandemics is evident by the large variance, within and between models, to predict the spread of COVID-19 [2]. This uncertainty misguides policies to mitigate epidemics and may result in catastrophic consequences such as financial crisis, massive unemployment, and the surge of the number of critically ill patients exceeding the capacity of hospitals. We've observed these consequences in the United States as a result of COVID-19. In particular, misunderstanding of risks and under/overestimation of efficacy of interventions can mislead policymakers about perception of evolving situations.

To improve prescriptive analytics for decision making, we propose Expert-in-the-Loop (EITL) prescriptive analytics using mobility intervention for epidemics. The key advantage of our approach is to leverage experts' best knowledge in estimating disease spreading and the efficacy of interventions which allows us to efficiently narrow down factors and the scope of combinatorial possible worlds. To manage uncertainty [8], we introduce our experience to develop simulations during the *Challenge on Mobility Intervention for Epidemics* [1]. We demonstrate that misconceptions about the causality can be corrected in the iterations of consulting with experts, developing simulations, and experimentation.

2 EXPERT-IN-THE-LOOP FRAMEWORK

In this section, we introduce our EITL framework, the goal of which is to (1) discover or update ground truths; (2) evaluate the efficiency and efficacy of each intervention in different situations; and (3) reduce combinatorial search space for optimization efficiently.

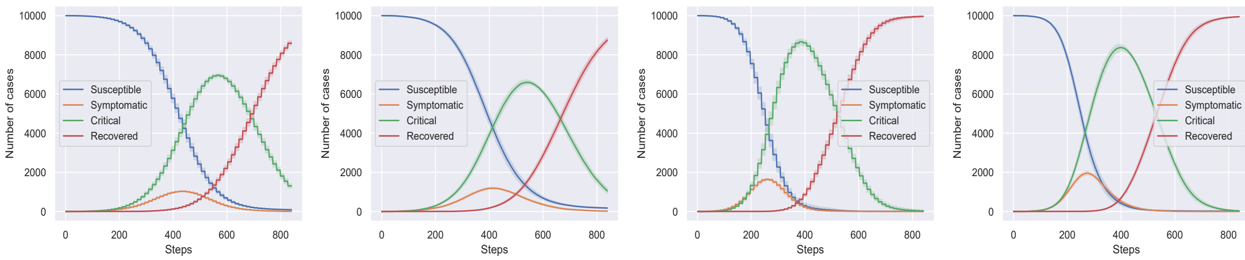
The motivation of our EITL stems from an hands-on experience of development of agent-based epidemic simulations [5, 6] and familiarity of the challenge designers' perspective against challengers[1]. Similar to the *Challenge on Mobility Intervention for Epidemics*, as a challenge designer, we have provided a black-box model that allows challengers to obtain only observable information and conduct experiments to find prescriptions. Due to complexity of social interactions in the simulation, it is a great obstacle for the challengers to discover ground truths.

In line with our formal experience, we aim at discovering ground truths prior to the outset. We explored multiple directions to understand the feasibility of each approach−given limited resources−to mitigate the simulated pandemic while also minimizing the cost of prescriptions. These approaches included genetic algorithms (GA), genetic programming (GP), reinforcement learning (RL), and rule-based heuristics. Each technique has pros and cons and best practices are to select methods that fit a specific task in the process.

[1]Geo-social simulation project web site: https://geosocial.joonseok.org

PAPW '20, August 24, 2020, San Diego, CA, USA

ACM ISBN 978-1-4503-6909-1/19/11...$15.00
https://doi.org/10.1145/3347146.3363350



(a) Ground Truth (Scenario 1) **(b) Estimated model (Scenario 1)** **(c) Ground Truth (Scenario 2)** **(d) Estimated model (Scenario 2)**

Figure 1: Comparison of epidemic dynamics between ground truth and calibrated model

For EITL, we repeat the following four steps: 1) to develop heuristics/rules to prescribe actions to agents, 2) optimize parameters of these rules to minimize evaluation score, 3) evaluate patterns within the results by consulting experts to understand resulting patterns and interpret causality, and 4) re-adjust the rules following our understanding and intuition.

Since it is difficult to control all factors that may influence outcomes, misconceptions about public health are common. In 1854, until John Snow [7], for example, traced the source of a cholera epidemic to a specific water pump, people regarded cholera as airborne epidemic.

To correct such a bias to the simulated world, we utilize modeling and simulation (M&S) to model the system as predictive and prescriptive analytics. Unlike black-box machine learning techniques, modeling and simulation provides deep understanding of the interested system. It is worth noting that the process of M&S allows experts to elaborate their knowledge and validate their theory and reasoning.

3 APPLICATION

The rest of the paper describes a best practice of applying EITL to *Challenge on Mobility Intervention for Epidemics* [1] and results.

3.1 Feasibility Study

In the real world, an expert group in response to epidemics may consist of epidemiologists, clinicians, social scientists, policy makers, health economists, community representatives and experienced simulation modelers. Its composition depends on disease types and society. In our settings, we leverage the simulator documentation [1] as the best knowledge including human mobility patterns, a disease model, feasible interventions, and its costs, obtained by the expert group consisting of epidemiologists, geoinformation scientists, social scientists, health economists, and policy makers. To validate our knowledge or ground truths, we develop micro-simulations that mimic the epidemic simulator. The main reason of adopting micro-simulations, instead of agent-based simulations [4], in this case is because we aim at statistical validation, and micro-simulations are sufficient to achieve our goal. The model includes the process of instanciating patient zeros who inherit the disease from the system without any contacts with other infectious individuals.

To understand the ground truth disease model, we conducted experiments by confining the population such that all agents stay in their neighborhood. After many iterations of modeling and calibration, we could find a disease model that behaves similar to the ground truth model. Figure 1 shows comparison of disease progression between the ground truth model and our estimated model in two different scenarios (see [1] for more information). It is worth noting that any subtle changes in the parameter configuration such as distribution of family, disease transmission period, or recovery period influence the shape of the curves. The process provides us a handle to control diseases with accurate estimation. If the model and parameters are inaccurate, interventions are ineffective having both huge false positive and false negative cases. For example, if we overestimate the probability of infection, then we are more likely to isolate or quarantine unnecessarily which leads to high intervention costs.

We categorize cases of transmission of epidemic into three: acquaintance contacts, stranger contacts, and unknown (also known as a patient zero). Given a probability P_c for an individual to get infected from an infected acquaintance contact and a probability P_s for an individual to get infected from an infected stranger contact, the probability P that individual x gets infected is:

$$P(x) = 1 - \prod_{i=1}^{t} (1 - P_c \cdot a_i) \cdot (1 - P_s \cdot (a_i + s_i)/n_i), \qquad (1)$$

where a_i is the number of acquaintance contacts with infectious individual at time i, s_i is the number of stranger contacts with infectious individuals at time i, n_i is the number of individuals where x is located at time i. We found that randomly selected patient zeros in Scenarios 1, 3, and 5 appear following the Poisson distribution where $\lambda \approx 2.4$ person a day.

Based on experimentation, we discovered the efficacy of four intervention types (confinement, quarantine, isolation, and hospitalization).

- *Confinement*: Confined individuals are allowed to interact with others in the same neighborhood including acquaintances. It mitigates at some level reducing major epidemic spreads from working areas. Acquaintance contacts are the main contribution and the number of acquaintance contacts at work is much larger than residential areas. This option is efficient and effective when the number of infections drastically increases (Scenario 2) or the number of infections is large (Scenario 4).
- *Quarantine*: Initially, experts assumed quarantined individuals contact only their acquaintances staying at home. However, by

experimentation we found that diseases from self-quarantined patients are transmitted to a stranger, which is plausible in the real world. We also discovered its efficacy is similar to confinement, but it is more expensive. This empirical findings is important for policymakers to make right decisions. For such a reason, our prescriptions do not use quarantine.

- *Isolation*: Empirically, we found the self-recovery period is distributed from 15 to 30 days including the pre-symptomatic period. To isolate infectious individuals is the most effective option to stop spreading from them since no contacts are allowed with the isolated individuals.
- *Hospitalization*: Similar to isolation, the hospitalized patients are not allowed to contact others. Since there is no penalty on critical cases nor reward on recovered cases, it has no merit of hospitalization, which is twice expensive than isolation. Therefore, we exclude hospitalization from interventions.

We take advantages of this knowledge to reduce the combinatorial search space for optimization.

3.2 Mitigation Strategies

Leveraging these observations, we optimize the combination of two types of interventions, namely confinement and isolation. To decide which agents to confine/isolate, we compute an infection probability (see Eq. 1) and a risk factor for each agent. A risk factor of an agent is a measure of how dangerous an individual would be if they were pre-symptomatic in such a way that the expected number of infections is estimated using their number of acquaintances and daily number of co-located agents. We categorize Scenarios 1, 3, and 5 into the same group using the same heuristics/rules due to its similarity. We highlight mitigation strategies for Scenarios 2 and 4 while strategies for other scenarios can be explained under the following general rules[2].

3.2.1 General rules. While infections from acquaintance contacts and stranger contacts can be inferred by contact tracing, there is no clue to infer who is a patient zero until they are discovered. Thus, whenever a new case is discovered, we isolate the symptomatic individual and select susceptible individuals with the high infection probability. There is the trade-off between the number of infections and the number of interventions. That is, if we isolate more individuals in question by decreasing a threshold σ, the number of pre-symptomatic individuals is more likely to decrease. If σ is too high, however, it ends up with more isolation because the number of infection cases increase. Therefore, it is the main challenge to find a perfect σ to make the balance between two measures.

Another dimension we take into account is a risk factor. Non-symptomatic individuals are isolated if the product of the probability and a risk factor exceeds a threshold ϵ. Suppose that there are two non-symptomatic individuals x and y having the same infection probability. If x has more acquaintances than y, then the risk of not isolating x is higher than that of y. Therefore, x is more likely to be isolated than y.

3.2.2 Scenario 2. The pandemic of Scenario 2 has higher infection rates where $P_c = 0.05$ and $P_s = 0.01$. Due to high infection rates, we need to isolate more aggressively, lowering threshold σ. However, controlling only σ is not sufficient to mitigate the spread of the pandemic due to high reproduction number. Our tactic for this scenario is to prevent gatherings with eleven or more people. If an individual has $n > 10$ acquaintances in a working area, we confine the individual with a probability $(n-11)/n$. This strategy can be seen as a work-from-home order.

3.2.3 Scenario 4. Scenario 4 is the extreme case that shows a trade-off between the number of accumulated infections and intervention costs. At the beginning, 300 pre-symptomatic are spawned and more than 100 susceptible individuals get infected from them during the first day. About 100-200 symptomatic cases including about 60 patient zeros are discovered the very next day. Therefore, about 240 patient zeros are unknown and our goal is to cost-effectively suppress initial spreading. Along with potential cases, we randomly select many individuals with high risk factors to isolate. Similar to Scenario 2 we prevent gatherings having more than five people in the working areas. We release them to reduce costs and the next day we repeat this process to find more pre-symptomatic cases.

4 CONCLUSION

In this paper, we introduced the Expert-in-the-Loop prescriptive analytics that leverages experts' best knowledge. While AI and machine learning can outperform in well defined problems such as optimization, experts can expedite the optimization process by means of modeling such a problem and choosing right tools. We also demonstrated how simulations can be used to discover ground truths as groundwork for prescriptive analytics. Although our mitigation strategies yielded fairly good results for general cases, we still have difficulty to narrow down the complex combination of rules for Scenario 2 and Scenario 4. In order to tackle the combination of rules, we will leverage evolutionary algorithms to explore diverse possible worlds for the future work.

ACKNOWLEDGMENT

This work was supported by the Defense Advanced Research Projects Agency (DARPA) under cooperative agreement No.HR00111820005 and the National Science Foundation Grant DEB-2030685. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

REFERENCES

[1] Challenge on mobility intervention for epidemics. https://prescriptive-analytics.github.io/challenge-cfp/index.html, Accessed July 24, 2020.
[2] Covid-19 ensemble dashboard. https://covid-ensemble.gmu.edu/, Accessed July 24, 2020.
[3] M. Gell-Mann. The interpretation of the new particles as displaced charge multiplets. *Il Nuovo Cimento (1955-1965)*, 4(2):848–866, 1956.
[4] J.-S. Kim, H. Jin, H. Kavak, O. C. Rouly, A. Crooks, D. Pfoser, C. Wenk, and A. Züfle. Location-based Social Network Data Generation Based on Patterns of Life. In *IEEE MDM 2020*, pages 158–167. IEEE, 2020.
[5] J.-S. Kim, H. Kavak, U. Manzoor, and A. Züfle. Advancing simulation experimentation capabilities with runtime interventions. In *SpringSim 2019*, pages 1–11. IEEE, 2019.
[6] J.-S. Kim, H. Kavak, C. O. Rouly, H. Jin, A. Crooks, D. Pfoser, C. Wenk, and A. Züfle. Location-based social simulation for prescriptive analytics of disease spread. *SIGSPATIAL Special*, 12(1):53–61, 2020.
[7] J. Snow. *On Mode of Communication of Cholera.* Churchill, London, England, 1855.
[8] A. Züfle, G. Trajcevski, D. Pfoser, and J.-S. Kim. Managing uncertainty in evolving geo-spatial data. In *IEEE MDM 2020*, pages 5–8. IEEE, 2020.

[2]Source code is available at https://github.com/joonseok-kim/kdd-papw20-challenge