Best subset selection of factors affecting influenza spread using bi-objective optimization

Aigerim Bogyrbayeva, Hanisha Tatapudi*, Hadi Charkhgard, Walter Silva-Sotillo

Department of Industrial and Management Systems Engineering, University of South Florida, Tampa, FL, 33620 USA

Abstract

A typical approach for computing an optimal strategy for non-pharmaceutical interventions during an influenza outbreak is based on statistical ANOVA. In this study, for the first time, we propose to use bi-objective mixed integer linear programming. Our approach employs an existing agent based simulation model and statistical design of experiments presented in Martinez and Das (2014) to generate the required data. Our computational results show that for an influenza outbreak with maximum attack rate of 33%, if we use the proposed optimal policy then the attack rate reduces to 0.6% with 85% accuracy. Our result significantly outperforms the existing result in the literature that estimates the influenza attack rate to be 1.83% under the proposed optimal policy.

Keywords: linear regression, best subset selection, multi-objective optimization, mixed integer linear programming

1. Introduction

Among the myriad of viral infections plaguing our surroundings, influenza is a common and highly contagious infection. A global outbreak of this virus is called an influenza pandemic, whose annual death estimates in the United States have increased from 3,349 in 1986 (Center for Disease Control and Prevention, 2010) to 56,000 in 2013 (Rolfes et al., 2016). In addition to the deaths, influenza has a significant economic burden on the society that it affects. Molinari et al. (2007) predicts the annual total economic burden of influenza epidemics to be $87.1 billion.

*Corresponding author
Email address: `tatapudi@mail.usf.edu` (Hanisha Tatapudi)

The methods used to contain this disease are broadly classified into pharmaceutical interventions (PIs) and non-pharmaceutical interventions (NPIs) (Qualls et al., 2017; Center For Disease Control and Prevention, 2007). PIs employed during an influenza pandemic are the administering of vaccines and antivirals. Although researchers argue that vaccine utilization is cost effective (see for instance Newall et al. (2010)), they are only about 60% effective in reducing the risk of influenza illness in the overall population (Treanor et al., 2012). In the absence of a vaccine or antiviral, the employment of NPIs is a cheaper alternative to mitigate an influenza pandemic and reduce mortality. NPIs employed during influenza are any activities that curb the spread of the disease such as quarantines, school or work closures and isolation (Center For Disease Control and Prevention, 2007; Hitoshi and et al., 2008). Analysis from earlier studies has given us confidence that NPIs are an inexpensive and efficient way for epidemic and pandemic mitigation (Bell et al., 2006; Lin et al., 2010; World Health Organization, 2012). Therefore, understanding the NPI factors that affect disease spread is critical to disease containment.

This study focuses on identifying the key NPI factors affecting an influenza spread using a new method. The main contributions of our research include (1) representing a linear regression model to predict Influenza Attack Rate (IAR) by selecting the best subset of NPIs using a bi-objective mixed integer linear program (BOMILP) and (2) establishing an optimal policy for NPIs using the proposed linear regression model. Note that IAR is a crucial parameter in measuring the severity of an influenza attack and represents the proportion of the population infected at the end of a pandemic.

The motivation behind our research is to empower decision makers with the flexibility of choosing the right mitigation strategies during an influenza outbreak. Often times, policy makers struggle from making successful preparedness plans due to inaccurate estimation of disease severity and lack of sufficient data. This was highlighted during the 2009 H1N1 outbreak when policy makers were accused of exaggerating the seriousness of the 2009 H1N1 pandemic (Doshi, 2011). Reports had identified that the initially predicted attack rate of 50% (Center for Disease Control and Prevention, 2010; Rolfes et al., 2016; Molinari et al., 2007) was higher than the actual 11% to 21%. This emphasizes a strong need for prediction models evaluating the effectiveness of strategies for disease containment.

In light of the above, regression models can be particularly useful for prediction. One particular research that focuses on designing NPI strategies during pandemic outbreaks using such models is Martinez and Das (2014). In that research, the authors employ a statistical ANOVA based design approach to develop a regression model and obtain optimal NPI strategies. Moreover, due to the absence of real-time data, an Agent-Based (AB) simulation model is utilized to simulate the disease transmission in a population. It is worth mentioning that the results obtained by Martinez and Das (2014) bring to attention that NPIs, when used under an optimal strategy for disease containment, show great potential to lower the IAR for an influenza outbreak.

Although the existing method and our research aim to design optimal NPI strategies, they are significantly different in terms of methodology. The existing study uses a statistical ANOVA based approach for IAR prediction, while the current study uses a BOMILP to represent a linear regression model for IAR prediction. It is worth mentioning that our method uses the agent based simulation model and statistical Design Of Experiments (DOE) model presented in Martinez and Das (2014) to generate the required data. Our computational results of our study show that for a pandemic with maximum attack rate of 33%, the proposed BOMILP minimizes IAR to 0.6% with 85% accuracy under an optimal strategy which is significantly better than the existing method's result, i.e., IAR is estimated to be 1.83% under the proposed optimal strategy by the existing method. Finally, we note that the BOMILP used in this study is proposed recently by Charkhgard and Eshargh (2017), and we are the first authors using that for predicting the spread of an infection.

The rest of the paper is organized as follows. In Section 2, we present the existing design of NPI strategies for pandemics. In Section 3, we discuss the general framework and model for the proposed BOMILP. In Section 4, we provide the computational results of the proposed method. Finally, in Section 5, we discuss some concluding remarks.

2. An existing method

In this section, we describe the existing method designed to develop NPI strategies for influenza containment presented in Martinez and Das (2014). This research simulates an influenza outbreak using AB simulation to generate disease transmission data for statistical DOE. The results from these statistical experiments were later used to develop optimal NPI strategies.

2.1. Agent based simulation

Simulation and modeling techniques have been instrumental in many studies related to epidemiology and disease transmission patterns. In fact many researches have utilized them to mimic transmission or contact processes, understand effects of viral outbreaks on a large scale and devise containment strategies (Glass et al., 2006; Roche et al., 2011; Bruce et al., 2010; Hitoshi and et al., 2008). AB simulation is a favorable choice for such case studies because it introduces granularity to the model by incorporating variables such as age groups, household size, health condition and disease status into the simulation. Due to the lack of real-time influenza data, an AB model that includes the above variables closely mimics the original state of the disease and appropriately replicates its effects on a population.

The idea behind using an AB model in the research presented by Martinez and Das (2014) was to simulate a living group of entities similar to a real population, track their daily activities and develop a detailed model of disease transmission for each individual. The model was created using a set of attributes that defines a population's characteristics such as age, gender, household, workplaces and schools along with health characteristics and weekday and weekend schedules, allowing the model to monitor the daily and hourly activities of individuals. The most important feature of the model is its capacity to track the hourly health status of each entity and measure the force of infection for each individual at the end of each day. This information is later used in a probabilistic model called the "infection transmission model" (Martinez and Das, 2014) to determine how susceptible an individual is to contract infection. This process is repeated for each person in the population, every day, till the end of the influenza period.

An interesting aspect of this model is the schedule set for these individuals. A predetermined schedule is assigned to them based on their age, gender and work status which controls a persons mixing groups by the hour. These mixing groups are simply other infected or susceptible contacts in the vicinity of an individual's workplace, school or local community. From each group an individual interacts, they contract an amount of infection at the end of hour T represented by the viral load.

The output of the model presents us with the total number of infected at the end of a pandemic. This information is used to calculate the most important measure of an influenza pandemic, the IAR. The results of this model are used to determine the optimal NPI strategies using a fractional factorial experiment discussed next.

2.2. Design of experiments

DOE is a branch of statistics that runs a series of controlled experiments to identify the factors that most influence the response of a model. Martinez and Das (2014) developed fractional factorial experiments to determine optimal NPIs strategies with respect to IAR.

The research used an AB simulation model, discussed previously, which focused on four major NPI actions: case isolation, household quarantine, school closure and workplace closure. However, to improve the level of detail, a number of performance measures were used to achieve a total of 16 NPI factors with each factor at 2 levels: low and high. In particular, cases to close a class in a school represents a number of infected cases to close a class, classes to close a school represents a number of closed classes to close a school and school closure duration represents a period of time when a school is closed (Martinez and Das, 2014). A list of these factors can be found in Table 1.

Table 1: NPI and their parameters considered in fractional factorial experiments

Factor	Intervention	NPI(1)	NPI(2)
1	Global Threshold	10	10
2	Deployment delay	3 days	7 days
3	Case isolation threshold	1 day	1 day
4	Case isolation duration	7 days	10 days
5	Case isolation compliance for workers	75%	75%
6	Case isolation compliance for non-workers	84%	57%
7	Household quarantine threshold	1 day	1 day
8	Household quarantine duration	7 days	7 days
9	Household quarantine compliance workers	75%	53%
10	Household quarantine compliance non-workers	84%	84%
11	Cases to close a class in a school	4	1
12	Classes to close a school	6	3
13	School closure duration	10 days	21 days
14	# cases to close a department in a workplace	6	3
15	% of departments to close a workplace	60%	30%
16	Workplace closure duration	10 days	7 days

Aiming to get a set of optimal NPI strategies, the model would have needed to run a set of 2^{16} experiments for a full evaluation of the factors. Additionally, replicates must be introduced into the model to account for randomness and increase the accuracy of prediction increasing the total to 327,680 runs, considering 5 replicates for each scenario with a fixed seed ranging between 1 and 5. To lessen this burden, a 2^{16-7} fractional factorial model was used, where out of 16 factors $(1/2)^7$ fraction of the number of possible factor level combinations were considered, which corresponded to the highest resolution. Fractional factorial experiments reduced the number of experimental runs from 65,536 to 512 per scenario, consequently reducing the total run time. Each run generates an IAR calculated as the proportion of influenza related deaths for a population. The results from the factorial design were used to develop a regression link with a statistical ANOVA based approach to find the significant NPI factors.

Martinez and Das (2014) concluded that school closure is the most significant NPI. The results show that increasing the number of cases to close a class results in a larger number of infected and increasing the school closure duration results in fewer number of infected in total. Based on these results, a regression model with R^2 value of 0.652 was developed. The regression model later was used to find optimal values for each NPI. The corresponding optimal IAR equal to 1.8% was found (for an outbreak with maximum attack rate of 33%).

3. A new method

The new method proposed in this section uses an implementation of a linear regression model to predict IAR using a multi-objective optimization technique. In linear regression, a model's bias and number of predictors play an important role in the accuracy of the prediction. Typically, modelers prefer achieving a desired accuracy with less bias and less number of predictors. However, compromising on one of the aspects, that is bias or the number of predictors, affects the outcome of the model. Hence, we look for an optimal trade-off between the both. Such a problem transforms into a bi-objective optimization problem with two objectives: minimizing bias and number of predictors. More precisely a BOMILP achieves these combined objectives and it can be stated as follows:

$$\min_{\hat{\boldsymbol{\beta}}\in\mathcal{F}} \{z_1(\hat{\boldsymbol{\beta}}), z_2(\hat{\boldsymbol{\beta}}\},$$

where $\mathcal{F}$ is the set of all feasible predictors, $z_1(\hat{\boldsymbol{\beta}})$ is bias, and $z_2(\hat{\boldsymbol{\beta}})$ is the total number of predictors.

Since an ideal solution that minimizes both objective function does not often exist, we are interested in understanding the trade-off between the two objectives. In other words, we are interested in computing the nondominated frontier which is defined next.

Definition 1. *A feasible solution $\boldsymbol{\beta} \in \mathcal{F}$ is called efficient or Pareto optimal, if there is no other $\boldsymbol{\beta}' \in \mathcal{F}$ such that $z_k(\boldsymbol{\beta}') \leq z_k(\boldsymbol{\beta})$ for $k = 1, 2$ and $\boldsymbol{z}(\boldsymbol{\beta}') \neq \boldsymbol{z}(\boldsymbol{\beta})$. If $\boldsymbol{\beta}$ is efficient then $\boldsymbol{z}(\boldsymbol{\beta})$ is called a nondominated point. The set of all nondominated points is referred to as the nondominated frontier.*

An illustration of the dominated and nondominated points in the objective space are shown in Figure 1.

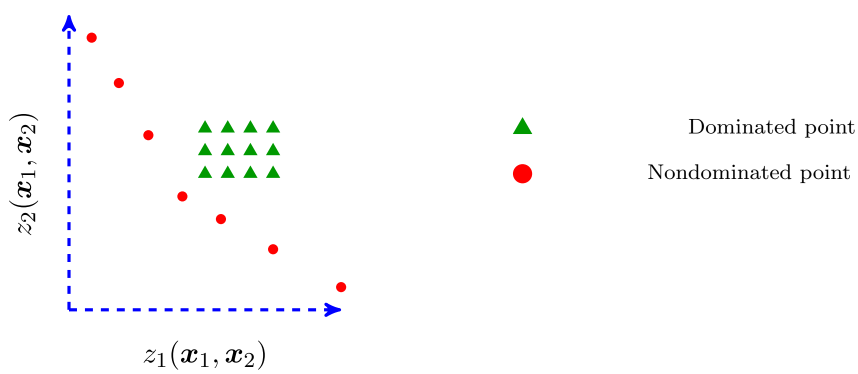


Figure 1: An illustration of the dominated and nondominated points in the objective space

We use the following BOMILP for selecting key NPI strategies to contain influenza spread which is based on the study of Charkhgard and Eshargh (2017). This model uses an absolute value of the sum of the residuals, which is the bias to formulate a linear regression as an integer linear program.

Let $X = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p] \in \mathbb{R}^{n \times p}$ be the model matrix, $\boldsymbol{\beta} \in \mathbb{R}^{p \times 1}$ be the vector of regression coefficients, and $\boldsymbol{y} \in \mathbb{R}^{n \times 1}$ be the response vector. The parameters of the model are the βs which are termed as the unknown variables and are estimated by $\hat{\boldsymbol{\beta}} \in \mathbb{R}^{p \times 1}$. To compute the best subset selection, we solve the following BOMILP, i.e.,

$$\min \sum_{i=1}^{n} \gamma_i \tag{1}$$

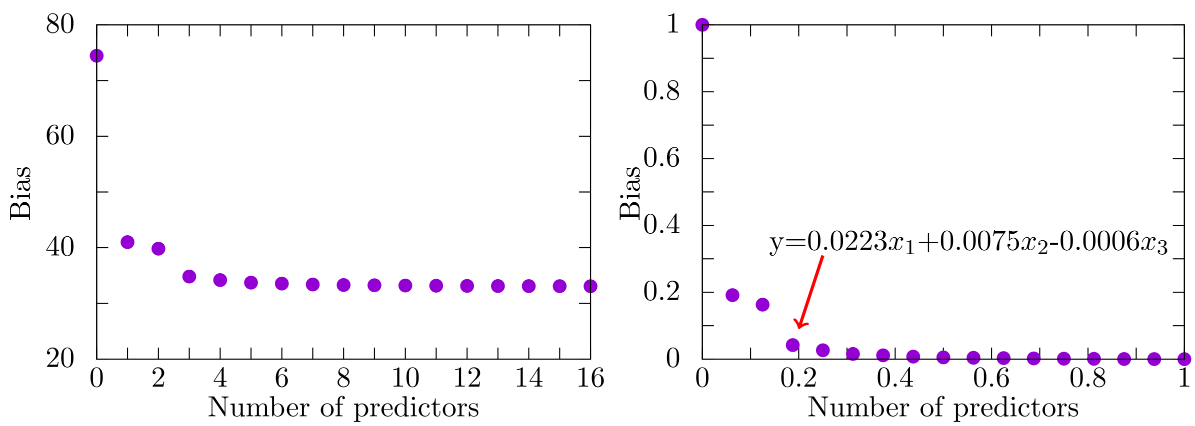
$$
\begin{aligned}
\min\ & \sum_{j=1}^{p} z_j && && (2)\\
\text{s.t. } & z_j l_j \leq \hat{\beta}_j \leq z_j u_j && \text{for } j = 1, \dots, p && (3)\\
& y_i - \sum_{j=1}^{p} x_{ij}\hat{\beta}_j \leq \gamma_i && \text{for } i = 1, \dots, n && (4)\\
& \sum_{j=1}^{p} x_{ij}\hat{\beta}_j - y_i \leq \gamma_i && \text{for } i = 1, \dots, n && (5)\\
& z_j \in \{0, 1\} && \text{for } j = 1, \dots, p && (6)\\
& \gamma_i \geq 0 && \text{for } i = 1, \dots, n && (7)\\
& \hat{\beta}_j \in \mathbb{R} && \text{for } j = 1, \dots, p, && (8)
\end{aligned}
$$

where $l_j \in \mathbb{R}$ and $u_j \in \mathbb{R}$ are a lower bound and an upper bound known for $\hat{\beta}_j$ for $j = 1, \dots, p$, γ_i for $i = 1, \dots, n$ is a non-negative continuous variable that captures the value of $|y_i - \sum_{j=1}^{p} x_j^i \hat{\beta}_j|$ in any efficient solution, and z_j for $j = 1, \dots, p$ is a binary decision variable that takes the value of one if $\hat{\beta}_j \neq 0$, implying that the predictor j is active. By these definitions, for any efficient solution, the first objective function takes the value of the sum of absolute residuals, and the second objective function computes the number of predictors. Constraint (3) ensures that if $\hat{\beta}_j \neq 0$ then $z_j = 1$ for $j = 1, \dots, p$. Note that since we minimize both objective functions (1) and (2) simultaneously, we must have that $z_j = 0$ if $\hat{\beta}_j = 0$ for some $j \in \{1, \dots, p\}$ in an efficient solution. Constraints (4) and (5) guarantee that $|y_i - \sum_{j=1}^{p} x_j^i \hat{\beta}_j| \leq \gamma_i$ for $i = 1, \dots, n$. Again note that since we minimize both objective functions (1) and (2) simultaneously, we must have that $|y_i - \sum_{j=1}^{p} x_j^i \hat{\beta}_j| = \gamma_i$ for $i = 1, \dots, n$ in an efficient solution.

4. Computational results

In this section, we present the computational results for best subset selection of NPIs. We set the maximum IAR rate to 33% in the AB simulation model. The motivation for the choice of 33% comes from the fact in 2009 H1N1 pandemic, less than one third of population was infected. To solve the BOMILP, we implement ε-constraint method in C++ and use CPLEX 12.7. All experiments are conduced on a Dell OptiPlex 7020 with 3.6GHz Intel Core i7 Processor, 16GB RAM with Windows 7 operating system.

For this computational study, we use the DOE results from Martinez and Das (2014), which consists of 2^{16-7} fractional factorial experiments as mentioned in Section 2. Before feeding this data to the BOMILP, we remove the outliers using InterQuartile Range (IQR), which is a difference between the 75th and 25th percentiles of the response data. All observations outside of the $[Q_1 - 1.5IQR, Q_3 + 1.5IQR]$ range are removed. In total, we use 2,143 observations to obtain the nondominated frontier of the proposed BOMILP.



(a) Not-normalized

(b) Normalized

Figure 2: Nondominated frontier

As illustrated in Figure 2a, the nondominated frontier consists of 17 points. As the bias decreases, the number of factors considered for prediction increases, thus it is impossible to reduce the bias without increasing the number of predictors. Therefore, choosing a nondominated point is not a trivial question. One method of selecting a nondominated point is to choose a point with a minimal Euclidean distance from the ideal point.

In our case, due to the large range of the values of the bias and the number of predictors, we normalize the nondominated points first (before selecting the desirable point). The standard normalization technique that we use translates the ideal point to the origin and ensures that the values of the bias and the number of predictors are between 0 and 1. Specifically, we use the following technique for normalization:

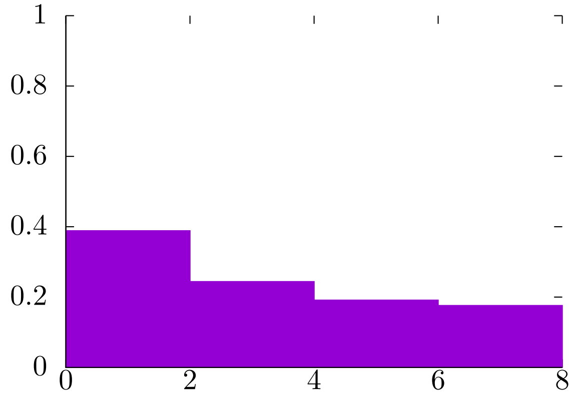
$$Normalized(e_i) = \frac{e_i - E_{min}}{E_{max} - E_{min}}$$

where, e_i is a value of the number of predictors (bias), E_{min} is the minimum value of the number of predictors (bias) and E_{max} is the maximum value of the number of predictors (bias) among nondominated points. Figure 2b shows the nondominated frontier after normalization and the chosen point based on the minimal distance from the ideal point (origin). The linear regression model obtained by our approach is $y = 0.0223x_1 + 0.0075x_2 - 0.0006x_3$ where x_1, x_2 and x_3 represent the cases to close class, cases to close school, and school closure duration, respectively.

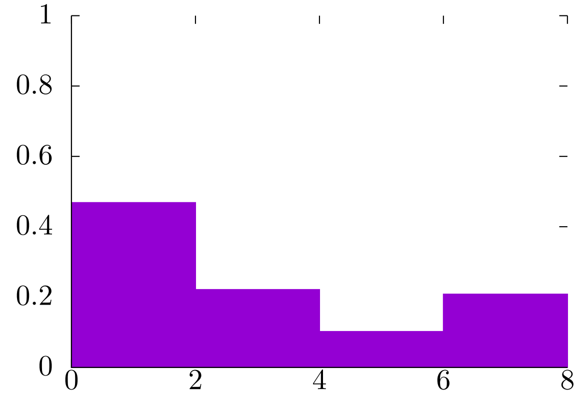
In other words, the obtained linear regression model suggests that three NPIs: cases to close class, cases to close school and school closure duration are the most important measures affecting influenza spread (for definitions see Section 2). Not surprisingly, this finding aligns with the conclusion of Martinez and Das (2014). Now, two natural questions arise: (1) How accurate is the proposed linear regression model? (2) How accurate would be the prediction of the proposed linear regression model under optimal policy?

To answer the first question, we run additional fractional factorial experiments. In particular, we run 2^{16-7} experiments with 5 replicates similar to Martinez and Das (2014) discussed in Section 2. However, we use random seeds ranging between 6 and 40 for AB model. Note, Martinez and Das (2014) used fixed seeds ranging between 1 and 5.

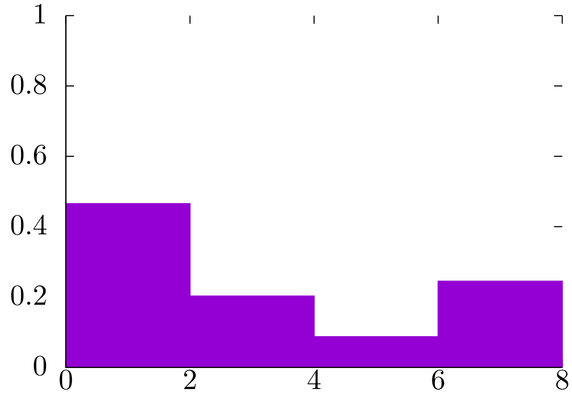
Figures 3 is the results obtained from assessing the accuracy of the linear regression. The figure shows the relative frequency of the absolute value of residuals for average of runs and individual runs. Outliers ae removed using IQR and residuals are calculated as the absolute difference between actual IAR and predicted IAR obtained based on the linear regression model. Our experimental results suggest that the proposed linear regression model predicts IAR values with more than 40% accuracy, while Martinez and Das (2014) obtained a linear regression model with $R^2 = 0.652$. Note that, the test experiments for evaluation of the linear regression are carried out on seeds between



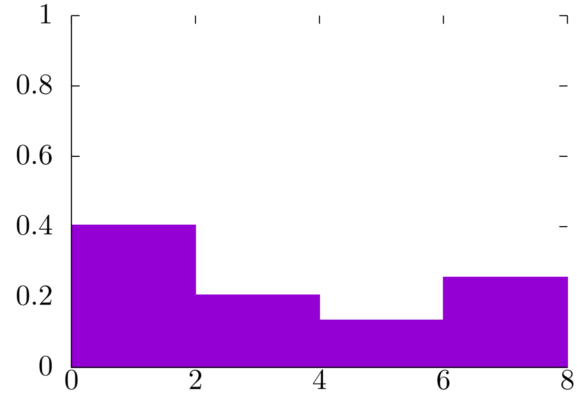
(a) Run 1



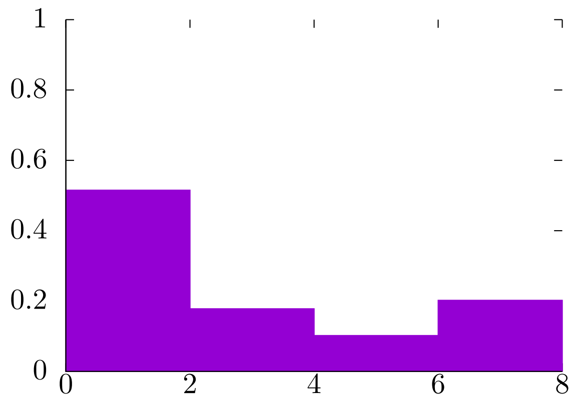
(b) Run 2



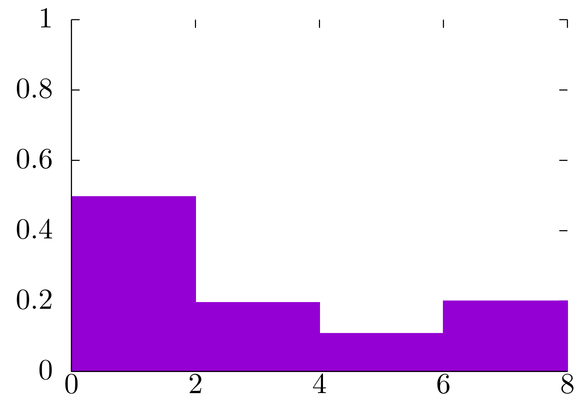
(c) Run 3



(d) Run 4



(e) Run 5



(f) Average

Figure 3: Relative frequency of residuals

6 and 40, while training is done on seeds between 1 and 5. So, one reason for the fact that the accuracy of our model is 40% is that we are using a different set of seeds for evaluation. Evidently, using the same seeds will increase the accuracy of our model (but our analysis would be less reliable then).

Now, to answer the second question, we first solve the following mathematical model to minimize the IAR, i.e., find optimal values of the key NPIs:

$$\min \sum_{i=1}^{p} \hat{\beta}_i x_i \tag{9}$$

$$\text{s.t. } v_i \leq x_i \leq w_i \qquad \text{for } i = 1, \ldots, p \tag{10}$$

$$x_i \in \mathbb{Z}_+ \qquad \text{for } i = 1, \ldots, p \tag{11}$$

where x_i is a positive integer value that captures the value of NPI $i \in \{1, \ldots, p\}$. Note that v_i and w_i are a lower bound and an upper bound for the value of NPI $i \in \{1, \ldots, p\}$ and are given in Table 1. Moreover, $\hat{\beta}_i$ is the coefficient of NPI $i \in \{1, \ldots, p\}$ in the obtained linear regression model. Note that an optimal solution of this optimization problem can be obtained easily by inspection. Basically by looking at the linear regression model, we can notice that cases to close class and cases to close school should take their maximum possible values while school closure duration should take its minimum possible value in order to minimize the objective function, i.e., IAR.

Specifically, we obtain the following optimal values of key NPIs: the number of cases to close class should be set to 1 case, the number of cases to close school should be set to 1 case, and school closure duration should be set to 42 days. The corresponding optimal IAR is 0.6% compared to 1.83% obtained by the DOE method as mentioned in Section 2. Thus, the result of the proposed approach significantly outperforms the result of obtained bu the DOE method.

Additional experiments are conducted on AB simulation model to validate the obtained result of the minimum IAR. In other words, we fix the values of key NPIs suggested by the proposed BOMILP to the obtained optimal values and let other NPI factors' values be random. In total, we run 2^{13-6} fractional factorial experiments with 5 replicates each and uses random seeds ranging between 6 and 40 to generate random data points. Figure 4 illustrates the minimum IAR validation experiments'

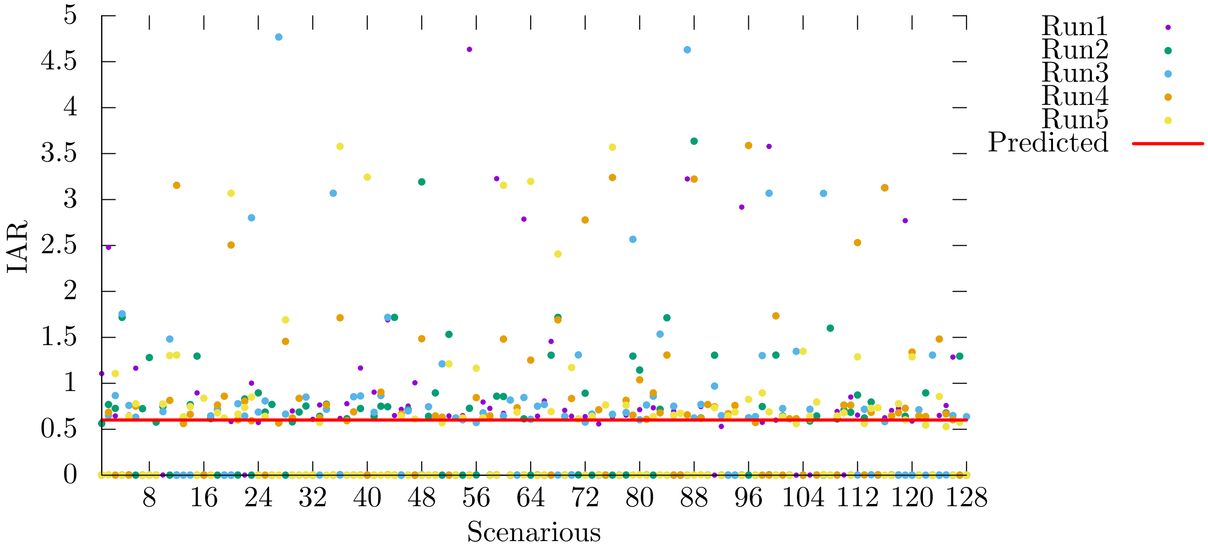


Figure 4: Actual vs. predicted IAR using random seeds

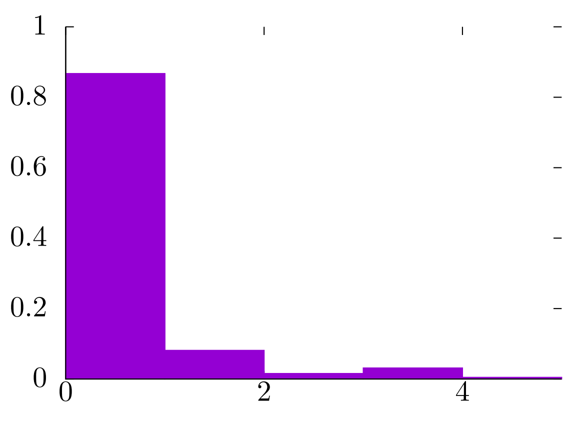


Figure 5: Relative frequency of IAR values under optimal policy

results, where the line indicates the minimum IAR value suggested by BOMILP model and the points represent the actual IAR values for 5 runs. As shown in Figure 5, the relative frequency of IAR under the proposed optimal NPI strategy indicates that 85% of IAR is between 0% and 1%. Thus, with 85% accuracy we can predict that the deployment of key NPIs at their suggested optimal values will reduce IAR value to 0.6%.

5. Conclusions

In this study, we used a bi-objective mixed integer linear program (BOMILP) to find the best subset of NPIs that affects the spread of influenza. It was shown that the proposed BOMILP has two advantages: (i) it computes a linear regression model consisting of the key NPI factors to predict IAR, (ii) it establishes an optimal NPI policy to minimize IAR. The proposed optimal policy of NPIs minimizes IAR to 0.6% with 85% accuracy (for an outbreak with maximum attack rate of 33%) which significantly outperforms the results of the existing method. We hope that the simplicity, versatility and performance of our approach encourage practitioners to consider using bi-objective optimization methods for predicting the spread of any infection and computing optimal policies.

References

Bell, D., Nicoll, A., Fukuda, K., et al., 2006. Nonpharmaceutical interventions for pandemic influenza, national and community measures. Emerg Infect Dis. 12 ((1): 88-94).

Bruce, Y. L., Brown., S. T., et al., 2010. A computer simulation of vaccine prioritization, allocation, and rationing during the 2009 h1n1 influenza pandemic. Vaccine 28 ((31): 4875-4879).

Center For Disease Control and Prevention, 2007. Interim pre-pandemic planning guidance: Community strategy for pandemic influenza Mitigation in the United States Early, Targeted, Layered Use of Nonpharmaceutical Interventions.

Center for Disease Control and Prevention, 2010. Estimates of Deaths Associated with Seasonal Influenza - United States, 1976-2007. Vol. 59.

Charkhgard, H., Eshargh, A., 2017. Best subset selection via bi-objective mixed integer linear programming. Optimization online.

Doshi, P., 2011. The elusive definition of pandemic influenza. Bulletin of the World Health Organization 89 (532-538).

Glass, R., Glass, J., Beyeler, L., et. al, 2006. Targeted social distancing designs for pandemic influenza. Emerg Infect Dis. 12 ((11): 1671-1681).

Hitoshi, O., et al., 2008. Major issues and challenges of influenza pandemic preparedness in developing countries. Emerg Infect Dis. 6 ((31): 875-880).

Lin, F., Muthuraman, K., Lawley, M., 2010. An optimal control theory approach to non-pharmaceutical interventions. BMC Infect Dis. 10 (32).

Martinez, D., Das, T., 2014. Design of non-pharmaceutical intervention strategies for pandemic influenza outbreaks. BMC Public Health 14 (1328).

Molinari, N.-A. M., Ortega-Sanchez, I. R., Messonnier, M. L., Thompson, W. W., Wortley, P. M., Weintraub, E., Bridges, C. B., 2007. The annual impact of seasonal influenza in the us: Measuring disease burden and costs. Vaccine 25 (27).

Newall, T. A., Wood, J., Oudin, N., MacIntyre, R., 2010. Cost-effectiveness of pharmaceutical-based pandemic influenza mitigation strategies. Emerg Infect Dis. 16 ((2): 224-230).

Qualls, N., Levitt, A., Kanade, N., et al., 2017. Community mitigation guidelines to prevent pandemic influenza united states, 2017. MMWR Recomm Rep. 66 (No. RR-1) (1-34).

Roche, B., Drake, J., Rohani, P., 2011. An agent-based model to study the epidemiological and evolutionary dynamics of influenza viruses. BMC Bioinformatics 12 (87).

Rolfes, M., Foppa, I., Garg, S., Flannery, B., Brammer, L., Singleton, J., et al., 2016. Estimated influenza illnesses, medical visits, hospitalizations, and deaths averted by vaccination in the united states.

Treanor, J., Talbot, H., Ohmit, S., et al., 2012. Effectiveness of seasonal influenza vaccines in the united states during a season with circulation of all three vaccine strains. Clin Infect Dis. 55 ((7): 951-9).

World Health Organization, 2012. Report of the WHO Pandemic Influenza A(H1N1) Vaccine Deployment Initiative.