Finding efficient intervention plans against Covid-19

Second place at the XPRIZE Pandemic Response Challenge

Nina Reščič[1,2,*,~], Vito Janko[1,*], David Susič[1,2,*], Carlo De Masi[1,*], Aljoša Vodopija[1,2,*], Matej Marinko[1,3,*], Tea Tušar[1], Erik Dovgan[1], Matej Cigale[1], Anton Gradišek[1], Matjaž Gams[1], Mitja Luštrek[1]

[1] Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia
[2] Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
[3] Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia
[*] Authors contributed equally
[~] Corresponding author: nina.rescic@ijs.si

***Abstract*— Covid-19 has so far affected every country in the world. The Non-Pharmaceutical Interventions (NPIs) by governments have proven themselves quite effective at stopping the spread of infections, but when applied in a very strict and long-lasting manner could have devastating consequences for the economic and social well-being of the population. XPRIZE and Cognizant organized the $500,000 XPRIZE Pandemic Response Challenge, where the participants were tasked to find good trade-offs between the costs and benefits of NPIs. This paper describes the solution by the team JSI vs Covid that placed second and won a $250,000 prize. The described solution uses an SEIR model to predict the spread of the infections, with the model parameters being dynamically changed based on active NPIs using machine learning. It then uses multi-objective optimization to find the desired trade-offs between NPI strictness and effectiveness.**

Keywords*— *Covid-19; countermeasures; Non-Pharmaceutical Interventions; epidemiological models; multi-objective optimisation

I. INTRODCUTION

The Covid-19 pandemic has negatively affected the whole world, with the virus spreading extremely fast. While non-pharmaceutical interventions (NPI) like closing schools and cancelling public events have proven effective at containing the pandemic [1, 2], they come with a large cost to the economy, and with a quality of life decrease for the general population. Policy-makers were thus given a challenging task of balancing the spread of the pandemic with the socio-economic costs of the NPIs. This task was made even harder due to how unprecedented the situation is and due to lack of reliable data about the exact extent to which the NPIs affect the spread of the virus.

As time goes on, however, more and more data about the pandemic becomes available (e.g., the number of infections in each country) and one could use artificial intelligence to 1) understand the effect of various NPIs, and 2) propose sensible intervention plans based on historical evidence and not just based on the intuition of policy makers. This was the exact idea of the XPRIZE: Pandemic Response Challenge [3] where two hundred research teams from all around the world competed to achieve the two previously mentioned tasks and to win the $500,000 prize purse (sponsored by Cognizant [4]). In this paper we describe our submission to this competition, how we tackled both problems and ultimately ended up as being one of the two winners [3].

First, in Section II we describe in more detail the two tasks given by the competition, then in Section III we describe our methods, and show the results in Section IV. Finally, we conclude in Section V.

II. THE PANDEMIC RESPONSE CHALLENGE

The competition was split into two phases. In the first one the "Prediction" phase the competitors had to predict the number of infections for 236 regions, given the NPIs that were in place in these regions (regions correspond to most countries in the world and some regions inside countries such as US states. To ensure fairness, the submitted models were tested each day after submission, for months, and were given actual NPIs in place at that time period and had to predict the number of infections from the submission date on. The models could use any other additional data if it was provided before the submission date.

In the second "Prescription" phase, the competitors had to create intervention plans for different situations (different countries and time periods) for two months in advance. There were 12 possible NPIs to pick from an OxCGRT database [5], each with different levels of strictness. An intervention plan could consist of any combination of these, and could change from day to day. For example, a possible intervention plan would be to use strict NPIs at the beginning, but gradually lower the stringency as time moves on and the predicted number of infections' fall. The prescribed intervention plans obviously could not be tested in real life so their quality was assessed based on two criteria, the predicted number of infections and the socio-economic cost. The prediction was made by the "standard predictor" provided by the organizers [6]. The socio-

economic cost of each NPI was provided by the organizers during the evaluation phase the submitted prescriptor was required to work with any cost provided. This mimics the real-life application of policy makers providing their own custom costs, fitted to the needs of their country. Each competitor could prescribe up to ten different intervention plans with different trade-offs between the two criteria. An intervention plan was considered better than another if it dominated it, which means that it was better on one criterion and not worse on the other. A solution (intervention plan) is said to be nondominated if there is no solution that dominates it. In a favourable case, the ten proposed plans should be spread all along the Pareto front - the image of all nondominated solutions.

III. METHODS

In Section III.A and Section III.B we describe our methods for the "Predictor" and "Prescription" phase of the competition, respectively.

A. Predictor

The goal of the "Predictor" phase was to predict the number of infections for each country/region for each day, months in advance, given the NPIs in that country/region. We did so by using a SEIR epidemiological model that was improved so that its "spreading rate" parameter β can dynamically adapt to the changes to the NPIs. The mapping between the NPIs and the spreading rate parameter was done using machine learning.

1) Datasets

We worked with two datasets. The first dataset consisted of daily reported infections and was used to fit the SEIR model (Section III.A.2). This data was collected from the Oxford Covid-19 Government Response Tracker (OxCGRT) database [5]. The second dataset was used for the β prediction model. While the main factors affecting the speed of the spread are the active NPIs, we collected the data on all other conditions we believed could be affecting it. We started with a dataset of 93 static (one per country) features such as development, culture, health, etc., which were extracted in our previous research on Covid-19 [7]. We then added "dynamic" features that could change day-by-day, namely the weather (temperature, humidity, etc.) and different NPIs collected from the OxCGRT database.

All the data was used for fitting the epidemiological model (Section III.A.3), but only a subset of 108 countries for training the machine learning model (Section III.A.4). The inclusion criteria for a country to be part of the training set were: sufficient data for that country, and negative correlation between NPI stringency and number of infections. The latter condition was due to some countries having inadequate testing and thus inaccurate data.

2) SEIR epidemiological model

One of the most commonly used approaches to predict the number of new daily infections are the epidemiological models. We used the standard SEIR model that uses *Susceptible*, *Exposed*, *Infected* and *Removed* compartments (FIGURE **1**).

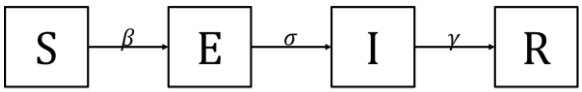

FIGURE 1: Scheme of SEIR model

The model uses parameters β, σ and γ that determine the transition probabilities from one (compartment) state to another as shown in the system below.

$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$
$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E$$
$$\frac{dI}{dt} = \sigma E - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

The β parameter (infection rate) was fitted based on the historical data using least square error, while the σ (incubation period) and γ (recovery rate) were set as static values based on the values found in the literature. The fitted β parameter would then be used as the prediction target for the machine learning model (Section III.A.3).

3) Fitting β

Since the βs are constantly (and sometimes drastically) changing over time in ways that cannot be modelled using a simple function, we decided to split the data from each region into intervals and fit them separately. These intervals were created in two different ways:

- depending on NPIs – if they changed by more than a predetermined threshold value a new interval is started, and
- depending on the infection trends – the intervals were created so that the number of infections were either rising or falling on each interval. The fitting was done separately on each set of intervals, and then we determined for each region which fitting gave better prediction accuracy and used those βs as the ground truth for the machine learning models in the next step.

4) Predicting β

Fitted βs from previous subsections were then used to create a machine learning problem where the goal was to predict βs from the features. To predict future infections, a sequence of βs were calculated from future data (e.g., NPIs) and then inserted into the SEIR model.

5) Other Components

The submitted pipeline works as follows:

1. If the NPIs in the given region have not changed after submission, take the last pre-calculated β for that region and use it in the SEIR model to predict the daily number of infections.
2. If the NPIs in the given region have changed, first use the features for that region (mainly the provided NPIs) in

conjunction with the created machine learning model to calculate the β parameter for each day. Then use calculated β in the SEIR model to predict the daily number of infections.

3. In all cases, when β changes the transition is made smoother by using exponentially weighted averages of the β values.
4. A linear model was used for β prediction. We performed feature selection with 153 collected features and those that showed by far the strongest correlation with β were the NPIs for the (t-14)-th day.

An exception to this procedure was made in roughly 40 regions where the SEIR model was not a good fit on the historical data, usually due to the too low number of daily infections. In these cases, we took the last month of known data for that region and then found other regions in the past that exhibited a similar infection pattern, i.e., their number of infections closely matched, when normalized for the population. We looked at what happened in those regions after the inspected period, and used this to predict the future in the original period. Since the predictions made in step 3 are expected to be of length less than 180 days, we used the standard procedure to create the remainder of the predictions.

B. Prescriptor

In the second competition phase we had to create intervention plans for different time periods and for different regions. Such intervention plans should have good trade-offs between the stringency of the interventions and the projected infections that result from them. Such problems are commonly tackled with multi-objective evolutionary algorithms (MOEAs) that imitate biological evolution to search the space of possible intervention plans, evaluate them in terms of their stringency and the number of infections, and find plans with good trade-offs between the objectives. We were facing a time constraint as well - we only had 6 hours to evaluate 235 regions (90 seconds on average per region). We used the NSGA-II [8] algorithm for the task. The intervention plans were represented as vectors, where the i-th variable represents what is the aggregated socio-economic costs of all NPIs to be used on the i-th week.

We decided to optimize with the granularity of one week instead of one day for two reasons: 1) it is unrealistic to expect real-life policies to change with a higher frequency and 2) the quality of the solutions did not substantially improve when using a smaller granularity. It is of note that this granularity parameter is adjustable if our system would be used in practice and a decision maker would so desire.

To evaluate such a vector of socio-economic costs during the optimisation process, they are first expanded so that each variable represents one day, then for each day the NPIs are selected so that they do not exceed the cost for that day and so that they are as effective in reducing the number of infections as possible. The effectiveness of every NPI combination according to the "standard predictor" was precomputed in advance, so that the selection in the previous step can be done with no computational overhead. Finally, the resulting matrix of NPIs for each day is sent to the standard predictor. The optimisation could in theory directly use the matrix representation where each value represents the presence (and strictness) of each NPI, but we have empirically evaluated that this only increases the search space and thus search time, without providing better solutions.

The resulting intervention plans, made by the described optimisation process, provided great trade-offs but the method turned out to be computationally too expensive, as each call to the standard predictor needed a few seconds for evaluation – and each region needed roughly 10,000 evaluations for the optimisation process to converge. Given the time constraint this process was much too slow.

We thus developed two methods derived from this standard multi-objective optimisation approach, and then combined them at the end.

1) Pre-computed plans

Our first method was to compute several plans in advance, and then for a specific region during the competition evaluation select the plan that is the most appropriate for the current situation in that region. The criteria for being "most appropriate" were the following: the desired length of prescription (we pre-calculated plans for 90-, 75-, 60- and 45-days, and selected the one closest to the desired length), infection trend (infections raising, falling, stable, raising fast, falling fast), and size of the country/region (small, large). For each of the listed combinations, ten prescriptions were pre-calculated and could be used for a given region during the competition. Since the socio-economic costs were still unknown at the time of the submission, our pre-calculated intervention plans only specified the maximum socio-economic costs for each day – which is in any case the natural representation of our optimisation process. Then during the evaluation, when actual costs for each NPIs were given, we selected the most effective NPIs as previously described.

2) Optimisation with the SEIR model

The second method used similar optimisation, but with two exceptions: 1) this optimisation was not done in advance, but directly for the country/regions and time intervals of interest, and 2) a fast surrogate model was used instead of the standard predictor. Surrogate model is a technique often used in optimisation where a computationally expensive model (in our case the standard predictor) is replaced by a simpler model that still returns similar results but is much faster. In our case the surrogate model was the same as epidemiological SEIR described in Section III.B with two differences. First, its parameters were fitted to the standard predictor's outputs instead of ground-truth infections. Second, the code was rewritten in Cython (static compiler for Python) to be faster. It still used the same pipeline of first using NPIs to determine the β parameter for each day and then dynamically changing that parameter during the SEIR evaluation.

3) Full prescriptor pipeline

The submitted pipeline works as follows:

1. For each region we first retrieved the data for the three weeks leading to the start date. This data is either stored in a historical file or is computed with the standard predictor if data is in the future. Based on this data, the country and prescription length, we chose the pre-computed plans described in Section III.B.1.
2. In the edge case where infection data always equals 0, we prescribe no interventions. Otherwise, we also run the optimisation described in Section III.B.2 to create new prescription plans from scratch.
3. Both pre-computed and surrogate optimisation return intervention plans of comparable quality, each having their strengths and weaknesses. The latter uses a surrogate model instead of the "real" one, and is done with severe time constraints, while the former optimizes for different (although similar) time/region combinations than the target ones.

We discovered, however, that combining both results frequently increases the overall quality of the obtained Pareto front approximation. Having the twenty solutions – ten from each method – we finally select ten best ones for the final submission. This step was done using the greedy Hypervolume Subset Selection (gHSS) method [9]. This approach finds an approximate solution to the hypervolume subset selection problem. In our case, the objective is to obtain the subset of ten solutions that maximize the hypervolume in the objective space. Large hypervolume values result in large dominated areas, therefore, solutions selected by gHSS are expected to dominate a large number of competitors' solutions.

IV. RESULTS

A. Predictor

1) Predicting β

Feature selection was run, however none of the features turned out to be selected more often than others. On the selected features we trained three models (linear regression, decision tree regressor and random forest regressor) with the target being *β*. Linear regression performed the best by far, but none of the models built on the selected features performed better than the model trained only on the NPIs. Thus, for the final model we selected linear regression and trained it only on the NPI data with the delay of (t – 14) days, since this delay turned out to have the strongest negative correlation with the normalized *β*s.

2) Competition Performance

The competition organizers calculated general mean average errors (MAE) and MAE by region to evaluate the predictors of every competitor. The full list of results can be found on [10], although the team names are anonymised. Our submission floated between fourth and first place, depending on the day of evaluation, and landed in second place on the last day of the evaluation, which meant we qualified for the second round. A sample prediction can be seen in FIGURE 2. The only consistent prediction error our method was doing (also visible on the same figure), was not taking holidays into account as the testing rate dropped significantly during such periods.

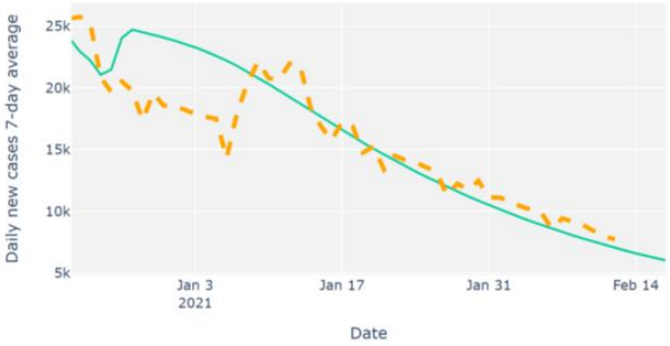

FIGURE 2: (Orange) The number of infections reported in Germany. The numbers were smoothed by using a 7-day moving average. (Green) The predicted number of infections in the same country/period.

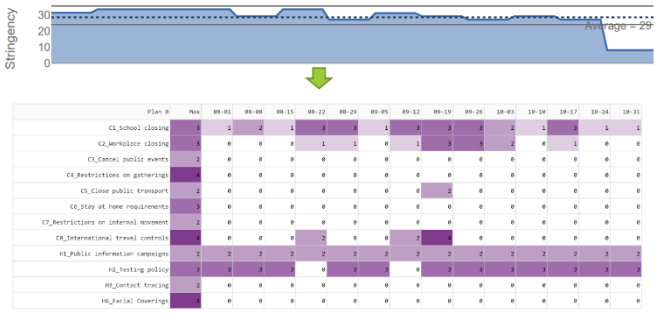

FIGURE 3: (Top) A prescribed plan for each week, where we list the maximum NPI cost for each week. (Bottom) A prescribed plan where each column represents one week, and each row is the intensity of a different NPI.

B. Prescriptor

Our system had to prescribe 10 intervention plans for each country/region for different time intervals and different socio-economic costs. The full list of results can be found at [11]. Sample prescriptions in the objective space are shown in FIGURE 3. The main criteria for the competition was the so-called domination count: a solution scored a point for each other solution it dominated. The points achieved by the top 10 teams are listed in TABLE 1. Numerically, we were the best performing team in the competition (while the numerical results were anonymized, the structure of our prescription was easily recognizable among the results) and when this was combined with the "qualitative score" of the judges, we landed in second place.

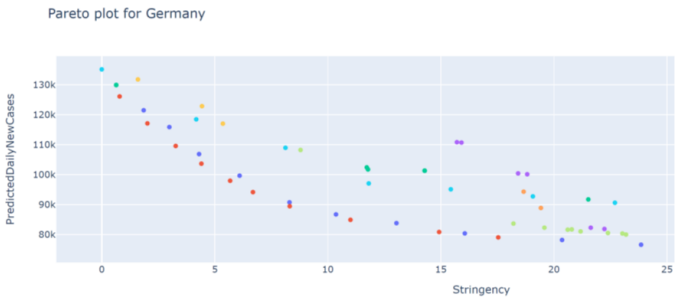

FIGURE 4: Prescribed plans from different teams for Germany. Each dot represents a different trade-off between the predicted number of cases and the NPI cost aggregated socio-economic cost. Some plans have low costs and a high number of infections, others vice versa. Our submission is represented by blue dots, and is visibly one of the two best ones.

TABLE 1: The domination count of the 10 best performing teams. Our submission is bolded.

Rank	Domination count
1	**515247**
2	490819
3	458146
4	435691
5	396968
6	313141
6	313141
8	288694
9	148766
10	134391

V. CONCLUSION

The XPRIZE: Pandemic Response Challenge focused on the development of data-driven AI systems to predict COVID-19 infection rates and prescriptions of intervention plans that regional governments, communities, and organizations can implement to minimize harm when reopening their economies.

While the problem of predicting new infections has been addressed many times, the real innovation of the competition was to find a way to prescribe NPI plans in such a way that both the number of infections and the stringency of the plans are the lowest possible. Our key insight when designing the predictor was to use machine learning to enhance the classical SEIR epidemiological model. This allowed us to dynamically adapt to the changes in NPIs as they were happening. On the other hand, the key insight for the prescriptor was to use MOEA methodology which is not common in this domain and then to find ways for making it less computationally expensive. The latter was done with a combination of surrogate model usage, computing sample prescriptions in advance, using weekly granularity for the optimisation and clever solution representation. Representing solutions using the overall stringency (rather than individual interventions) lead to far more effective optimisation (due to search-space reduction) and consequently better intervention plans in a reasonable time.

Another important issue to explore is how to present such methods and their outputs to decision-makers. We developed a prototype web application that could be used for such a purpose, but collaboration with actual decision-makers is necessary to test and improve it.

ACKNOWLEDGMENT

This research was funded by Slovenian Research Agency (research core funding No. P2-0209 (B)).

REFERENCES

[1] S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker and e. al., "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe," *Nature,* 2020.

[2] S. Moore, E. M. Hill, M. J. Tildesley, L. Dyson and M. J. Keeling, "Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study," *The Lancet Infectious Diseases,* vol. 21, no. 6, pp. 793-802, 2021.

[3] XPRIZE. [Online]. Available:https://www.xprize.org/challenge/pandemicresponse. [Accessed 20 June 2021].

[4] Cognizant. [Online]. Available: https://www.cognizant.com/. [Accessed 20 June 2021].

[5] T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar and H. Tatlow, "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)," *Nature Human Behaviour,* p. 529–538, 2021.

[6] R. Miikkulainen, O. Francon, E. Meyerson, X. Qiu, D. Sargent, E. Canzani and B. Hodjat, "From Prediction to Prescription: Evolutionary Optimization of Nonpharmaceutical Interventions in the COVID-19 Pandemic," *IEEE Transactions on Evolutionary Computation,* vol. 25, no. 2, pp. 386-401, 2021.

[7] V. Janko, G. Slapničar, E. Dovgan, N. Reščič, T. Kolenik, M. Gjoreski, M. Smerkol, M. Gams and M. Luštrek, "Machine Learning for Analyzing Non-countermeasure Factors Affecting Early Spread of COVID-19," *Preprints,* 2021.

[8] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation,* vol. 6, no. 2, pp. 182-197, 2002.

[9] A. P. Guerreiro, C. M. Fonseca and L. Paquete, "Greedy hypervolume subset selection in low dimensions," *Evolutionary Computation,* vol. 24, pp. 521-544, 2016.

[10] XPPRIZE, "XPRIZE Pandemic Response Challenge: Phase1 Results," [Online]. Available: https://phase1.xprize.evolution.ml/. [Accessed 20 June 2021].

[11] XPRIZE, "XPRIZE Pandemic Response Challenge: Phase2 Results," [Online]. Available: https://phase2.xprize.evolution.ml/. [Accessed 20 June 2021].