

The Nebraska Team Qualitative Report Phase 2

Dr. Fadi Alsalem, Assist. Prof. Architectural Engineering (UNL), Dr. Alison Freifeld, M.D. Int. Med. Infectious Diseases (UNMC)
Dr. Basheer Qolomany, Cyber-systems (UNK), Dr. Dan Piatkowski, Assist. Professor, Comm.& Regional Planning (UNL)
Mr. Mostafa Rafaie, Data Science, Mutual of Omaha, PhD student (UNL), Mr. Mohammad Ali Takallou, PhD student (UNL)
Mr. Ali Ramini, PhD student (UNL)

1. Introduction:

The Xprize pandemic response challenge is a four-month challenge that consists of two phases and focuses on the development of data-driven AI systems. The goal of the competition in phase one was to predict COVID-19 daily cases for each region accurately, and in the current phase, the target is to prescribe Intervention Plans (IPs) that regional governments, communities, and organizations can implement to minimize harm when reopening their economies. In other words, our focus in this part is to develop prescription models for any region that generate useful intervention plans to minimize the number of daily cases while minimizing the stringency of the recommended interventions based on a time sequence of the number of cases in a region and the past intervention plans in place. Our general model approach uses the well-known NeuroEvolution of Augmenting Topologies (NEAT) which is a genetic algorithm (GA) for the generation of evolving artificial neural networks. Adopting the NEAT algorithm turned our problem into an optimization problem. In this optimization, our goal is to develop a model for a maximum of 90 days with up to 10 different prescriptions for any regions that have the potential to minimize the number of daily cases while minimizing the stringency of the recommended interventions. Our main contribution in this phase, as explained in detail below, is to find the best fitness function that can be used to alter the weighting parameters and structures of networks while searching for the most optimum solution. Figure.1 shows the general approach for our model.

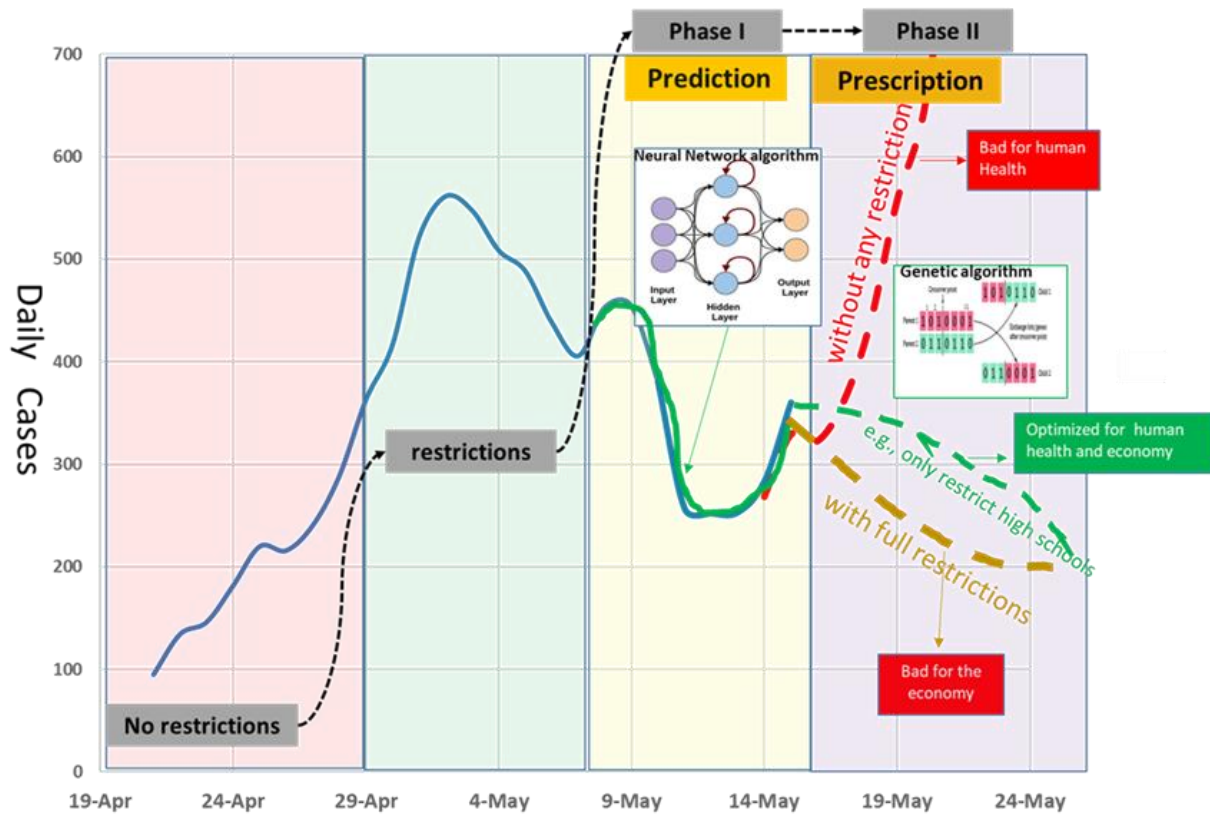


Figure 1: A schematic for the prediction/prescription modeling approach.

2. Approach detailed Explanation:

The main goal of phase 2 is to optimize the number of cases and stringency for a given region. The original fitness function is defined as $\text{fitness} = -(\text{new cases} * \text{stringency})$, with an optimization target to achieve a very small absolute value. A given model with a small fitness value indicates a better strategy to be adopted for a given region. In reinforcement learning, an iterative algorithm is said to converge when, as the iterations proceed, the output gets closer to some specific value. Next, we evaluated different fitness functions that achieve a target performance with less generation. As shown in Figure 2 we found the exponential

function is the most promising fitness function to achieve the best performance with fewer generations. Specifically, the new adopted fitness function is defined as $(e^{\text{new cases}} * e^{\text{stringency}})$. This novel adjustment we believe is working as it tunes the reward value exponentially to adopt for high fitness numbers and leans it faster to lower the number. However, while the solution starts to converge faster with this new fitness function, the different scales between the average case numbers and stringency

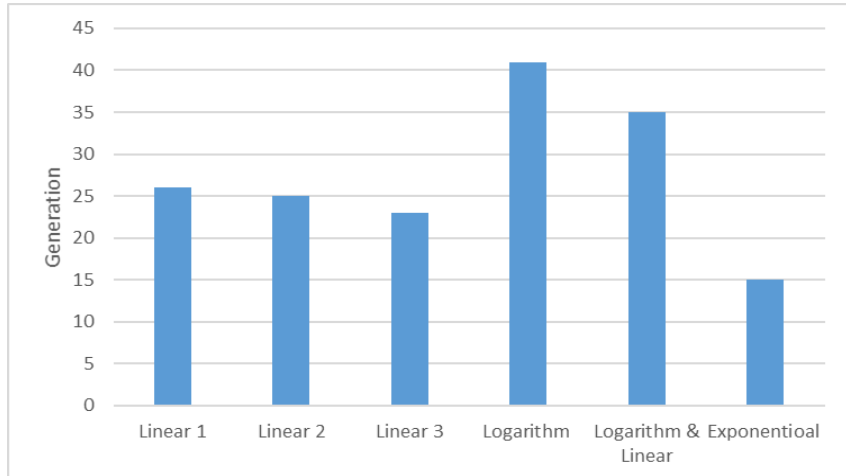


Figure 2: The convergence evaluation of different fitness functions

imposed a challenge for convergence to small numbers. To address this problem, we adopted a heuristic approach to scale the two parameters by different constant values as described by $\text{fitness} = (e^{\text{new cases}/2500} * e^{\text{stringency}/500})$. We also, according to our previous experience in phase 1, used 28 days for the number of look-back days during the training. Figure 3 shows the convergence of the fitness versus 300 generations for 9 out of 10 genomes. Moreover, the best fitness among different genomes in each generation that is adopted to provide the optimal solution is extracted and shown in Figure 4.

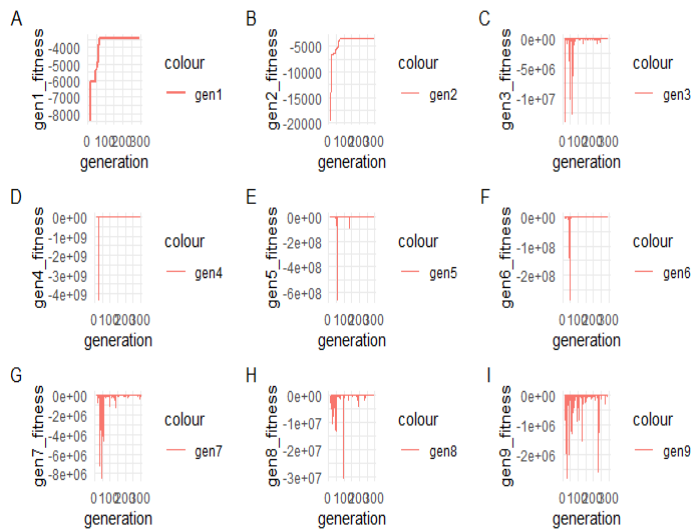


Figure 3: The trend of updating fitness variable versus 300 generations for 9 out of 10 genomes.

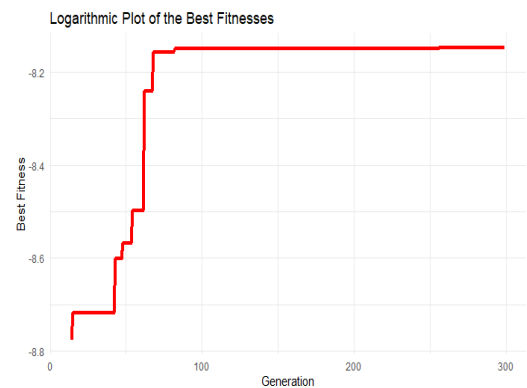


Figure 4: the best fitness between different 10 genomes in each of the 300 generation is extracted.

1. Innovation

Our innovation lies in the development of an optimized prescription model with recommendations to reduce the future COVID-19 cases while minimizing the impact of imposing the restriction. While we have explored multiple novel options, due to the limited time for implementation and space for explanation, our focus was on improving the optimization function of a NEAT model. Examples of other promising, but still in the development phase are:

- **Vacation days:** adapt the prescription model so it accounts for different vacation days for different countries
- **Vaccinations:** we believe vaccination is a very important factor. However, we decided not to include it to our current model, as its effect had not shown yet on the historical data we used to train the model
- **Unsupervised learning:** we are working on a novel data unsupervised labeling solution that can be used in the prediction and prescription process
- **The reproduction number (R):** as explained in phase 1 we plan to estimate the R-value for a given region in the US that can be used to tune both the prediction and prescription models using some wireless sensing technologies such as the Kinsa wireless thermometer (<https://kinsahealth.co/>).

2. Generality:

The developed Prescription model shows great performance across multiple different regions in the world.

3. Collaborative contributions:

Our general model approach uses the well-known NEAT approach with a modified fitness function, and the model will be open source and can be shared with others to benefit from.

4. Consistency:

Our model accuracy so far is very consistent for predicting future events. Similar behavior was observed when validating the model using the historical data.

5. Speed and resources use:

The base code supports a single process training. The code was redesigned to be able to utilize more process units in a time (parallel). The current server configuration is 16CPU, 24 GB RAM.

6. Addressing the challenge:

The model is developed using a holistic approach and is a very module. Thus, it can easily be adapted to accept new input such as the Kinsa thermometer data and account for other restrictions modes and new developments such as vaccinations and vacation days. However, due to the time limit, the focus of the current model is to account for the provided OXFORD resections modes.