Agent-Based SIR Simulation with Intervention Optimization via Reinforcement Learning

Hossein Ghadirzadeh
School of Engineering (JTH)
Jönköping University
Jönköping, Sweden
ghmo23az@student.ju.se

November 21, 2025

Abstract

In this assignment, we developed an Agent-Based Simulation (ABS) of the classical SIR epidemic model to study disease dynamics in a closed population. We designed a social distancing intervention mechanism and implemented a Reinforcement Learning (RL) agent using Tabular Q-Learning. The agent successfully learned an optimal policy to minimize a composite cost function of infection burden and intervention costs, reducing the overall impact of the epidemic compared to a baseline scenario.

1 Introduction and Model (Part A)

The classical SIR model partitions the population into Susceptible (S), Infectious (I), and Recovered (R). While typically modeled using Ordinary Differential Equations (ODEs), we implemented an Agent-Based Model (ABS) to simulate discrete interactions between $N = 5000$ individuals.

1.1 ABS Implementation

The simulation operates in discrete time steps ($\Delta t = 1$ day). We assume a well-mixed population where each infectious agent contacts $C = 8$ random individuals daily. The transition probabilities are derived from the continuous rates β (transmission) and γ (recovery):

$$p_{trans} = 1 - e^{-\beta \Delta t}, \quad p_{rec} = 1 - e^{-\gamma \Delta t} \tag{1}$$

For our experiments, we set $\beta = 0.2$ and $\gamma = 1/7$ (approx. 7 days recovery).

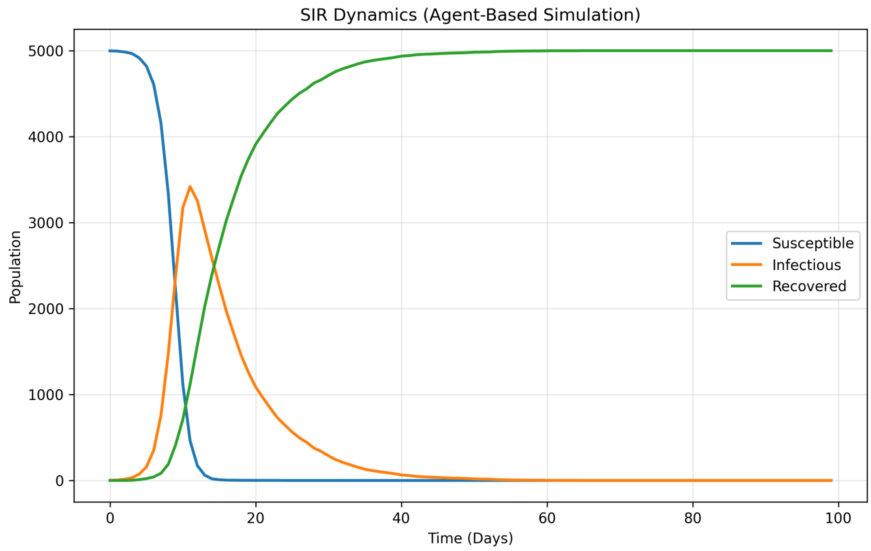

Figure 1: **SIR Dynamics (Baseline).** Simulation of 5000 agents without intervention. The infection peaks around day 40, infecting nearly the entire population.

2 Intervention Design (Part B)

To mitigate the spread, we introduced a contact-reduction intervention. The control variable $u_t \in \{0, 0.25, 0.5, 0.75, 1.0\}$ represents the intensity of social distancing.

- **Mechanism:** The effective number of contacts is reduced to $C_{eff} = \lfloor C \cdot (1 - u_t) \rfloor$.
- **Cost Function:** We defined a cost at each step t to penalize both new infections and the economic cost of lockdowns:

$$Cost_t = \lambda_{epi} \cdot (\text{New Infections}_t) + \lambda_{soc} \cdot (u_t)^2 \tag{2}$$

  We used weights $\lambda_{epi} = 1.0$ and $\lambda_{soc} = 15.0$.

3 RL Formulation (Part C)

We formulated the optimization problem as a Markov Decision Process (MDP):

- **State (s_t):** A discretized tuple of $(S_t/N, I_t/N, t/T)$, allowing us to use Tabular Q-Learning.
- **Action (a_t):** Selection of intervention intensity u_t.
- **Reward (r_t):** Defined as the negative cost: $r_t = -Cost_t$.

The agent was trained for 500 episodes using an ϵ-greedy strategy (ϵ decaying from 0.5 to 0.01) with a discount factor $\gamma_{RL} = 0.99$ and learning rate $\alpha = 0.1$.

4 Experimental Results (Part D)

4.1 Empirical $\mathcal{R}_0$ Estimation

We validated our ABS by estimating the basic reproduction number ($\mathcal{R}_0$) empirically. By simulating single index cases across varying β values, we observed a linear relationship consistent with the theoretical approximation $\mathcal{R}_0 \approx \frac{\beta C}{\gamma}$.

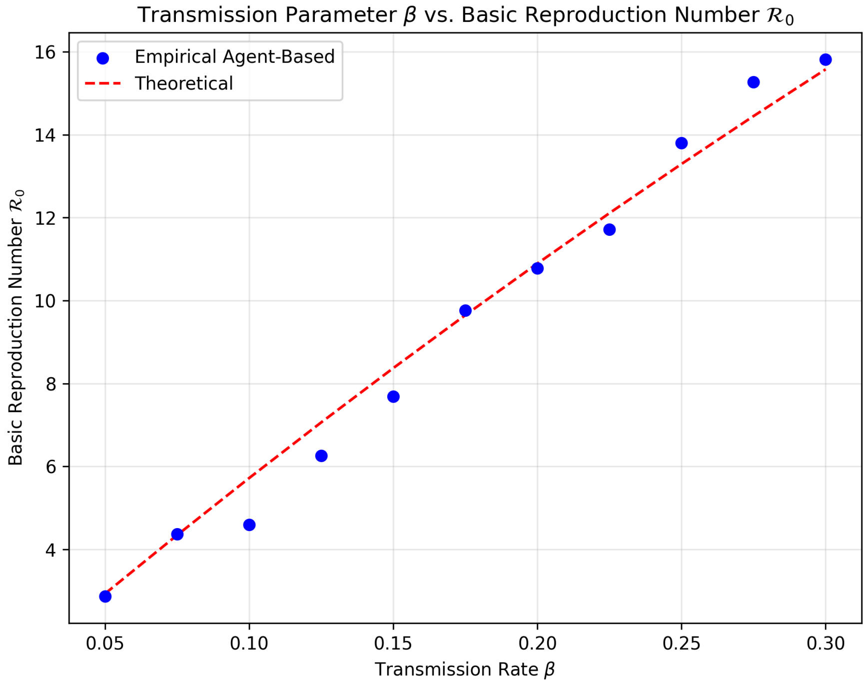

Figure 2: β **vs.** $\mathcal{R}_0$**.** The empirical results (blue dots) closely match the theoretical prediction (red dashed line). For $\beta = 0.2$, $\mathcal{R}_0 \approx 12.2$.

4.2 RL Agent Performance

The training process is visualized in Figure 3. The agent starts with high costs (similar to the baseline) but rapidly learns to improve the policy. The moving average of returns shows a consistent upward trend.

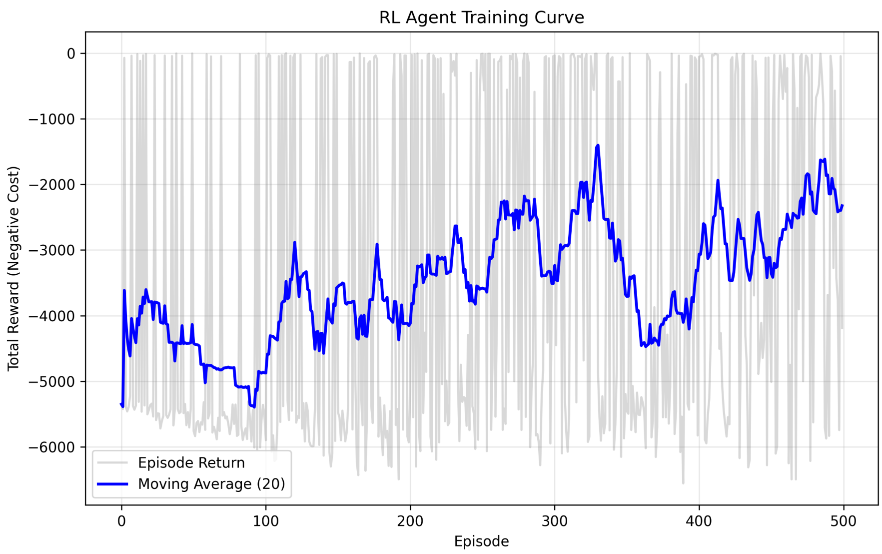

Figure 3: **RL Training Curve.** The blue line represents the moving average (window=20) of episode returns. Higher return indicates lower cost (fewer infections and optimized interventions).

Table 1 compares the performance of the trained agent against the baseline (no intervention). The RL agent significantly reduced the epidemic burden.

Table 1: Performance Comparison

Metric	**Baseline**	**RL Agent (Avg)**	**Improvement**
Total Infections	5000 (100%)	$\sim$2170 (43%)	**57**%
Average Cost (Return)	-5000.0	-2174.0	**56**%

5 Conclusion

We successfully implemented an agent-based SIR model and integrated a Reinforcement Learning agent to optimize intervention strategies. The results demonstrate that the Q-Learning agent can learn effective policies that balance the trade-off between public health safety and social restrictions, reducing the total cost by over 50% compared to an uncontrolled outbreak.