Mathematical Biosciences and Engineering

MBE, 20(8): 1–7.
DOI: 10.3934/mbe.2023640
Received: 10 October 2022
Revised: 10 June 2023
Accepted: 15 June 2023
Published: 29 June 2023

http://www.aimspress.com/journal/MBE

Supplementary material

Deep reinforcement learning framework for controlling infectious disease outbreaks in the context of multi-jurisdictions

Seyedeh Nazanin Khatami[1,*] and Chaitra Gopalappa[2]

[1] MGH Institute for Technology Assessment, Harvard Medical School, Boston, MA 02114, USA
[2] Mechanical and Industrial Engineering Department, University of Massachusetts Amherst, Amherst, MA 01003, USA

* **Correspondence:** Email: skhatami@mgh.harvard.edu.

S.1. Details of DQN algorithm

Throughout the training, a Q-table gets updated in the Q-learning algorithm, where each table element represents a state-action value. In the case of ample state space, when building a Q-table is intractable, or in the case of continuous state space, a Q-function is used to map state-action pair to a Q-value. Deep neural networks are used as a function approximator of state-action pair to a Q-value. During training, the algorithm learns the weights of the deep neural network. As a result, given the state to the deep Q-network (DQN), the Q-value associated with each action is outputted. And hence, the highest Q-value corresponds to the optimal action.

In the case of this study, the trained Q-network works as follow:

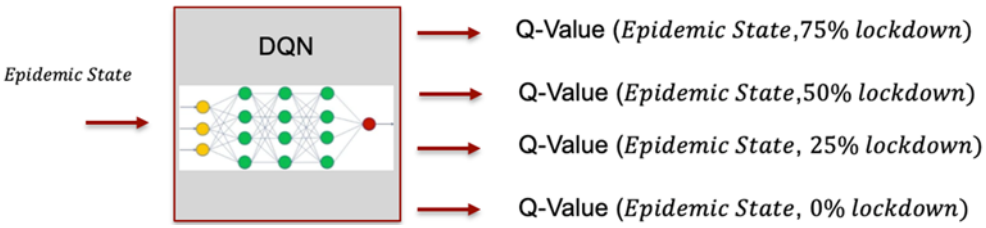

Figure A: Schematic of DQN in the context of this study.

Given the epidemic state, the Deep Q-network outputs the Q-value associate to every action. And hence, the highest Q-value corresponds to the optimal action in that epidemic state.

In the DQN algorithm, there are two networks (two artificial neural networks), and an experience replay:

- Q-Neural network: is usually a deep neural network,
- Target Neural Network: is identical to the Q-Neural network,

- Experience replay: is used to memorize the agent's experience when interacting with the environment, i.e. (state, action, next state, reward) as to reduce the correlation between the agent's experiences and prevent overfitting of the network.

DQN algorithm is described as bellow [1]:

- Initialize replay memory to some capacity.
- Initialize Q-network with random weights.
- For a pre-defined number of episodes:
  - At each training step, random samples from experience replay are selected.
  - Batch of training data is fed into the Q-network and target network.
  - An action is taken based on the pre-defined action selection strategy, i.e., epsilon greedy.
  - DQN and target network separately predict the Q-values of current state and all the actions.
  - The experience is stored as a pair of (state, action, next state, reward) in the replay buffer.
  - Mean squared loss gets calculated based on the Q-value of target network and Q-network.
  - The loss gets backpropagated to the Q-network so to update the weights using gradient descent algorithm.
  - After a pre-defined number of time-steps, Q-network weights get copied to the target network. Q-network and target network become identical again.

S.2. Summary of Optimal Policy for Scenarios 1 to 5

Table S1: Summary of scenarios 1 to 5. Shaded cells are the optimal policy in terms of when to be implemented (number of days of delays) and how to be implemented (number of days in each lockdown categories).

Delays			**30 days**	**45 days**	**60 days**	**75 days**	**90 days**
Scenario 1	Observed Prevalence (%)		0.002	0.01	0.054	0.272	1.359
	Actual Prevalence (%)		0.0032	0.016	0.085	0.43	2.134
	Number shutdown	0%	285	285	285	285	285
		25%	54	54	54	54	54
		50%	61	61	61	61	61
		75%	0	0	0	0	0
	Number of days of hospital capacity exceeded		0	0	0	0	0
Scenario 2 & 4	Observed Prevalence (%)		0.0023	0.0127	0.0676	0.353	1.805
	Actual Prevalence (%)		0.0037	0.02	0.1079	0.561	2.844

Continue to next page

Delays			**30 days**	**45 days**	**60 days**	**75 days**	**90 days**
Scenario 2	Number shutdown	0%	292	292	292	292	292
		25%	44	44	44	44	44
		50%	27	27	27	27	27
		75%	37	37	37	37	37
	Number of days of hospital capacity exceeded		0	0	0	0	0
Scenario 4	Number shutdown	0%	292	292	292	292	292
		25%	41	41	41	41	41
		50%	67	67	67	67	67
		75%	0	0	0	0	0
	Number of days of hospital capacity exceeded		0	0	0	0	0
Scenario 3 & 5	Observed Prevalence (%) Actual Prevalence (%)		0.0025 0.004	0.0139 0.0223	0.0741 0.118	0.384 0.610	1.946 3.057
Scenario 3	Number shutdown	0%	343	343	343	343	343
		25%	0	0	0	0	0
		50%	0	0	0	0	0
		75%	57	57	57	57	57
	Number of days of hospital capacity exceeded		0	0	0	0	0
Scenario 5	Number shutdown	0%	305	305	305	44	305
		25%	24	24	24	27	24
		50%	71	71	71	37	71
		75%	0	0	0	0	0
	Number of days of hospital capacity exceeded		0	0	0	0	0

Continue to next page

Delays			**95 days**	**100 days**	**105 days**	**110 days**	**120 days**
Scenario 1	Observed Prevalence (%) Actual Prevalence (%)		2.305 3.60	3.878 6.019	6.438 9.877	10.465 15.763	24.787 34.90
	Number shutdown	0%	292	300	306	308	368
		25%	46	27	36	41	1
		50%	62	73	52	34	11
		75%	0	0	6	17	20
	Number of days of hospital capacity exceeded		0	0	5	16	36
Scenario 2 & 4	Observed Prevalence (%) Actual Prevalence (%)		3.073 4.811	5.1716 8.016	8.544 13.035	13.725 20.432	30.86 42.29
Scenario 2	Number shutdown	0%	292	297	342	353	353
		25%	44	42	0	0	0
		50%	27	20	0	0	0
		75%	37	41	58	47	47
	Number of days of hospital capacity exceeded		0	0	11	24	58
Scenario 4	Number shutdown	0%	292	342	341	351	373
		25%	41	0	8	4	6
		50%	67	0	10	12	1
		75%	0	58	41	33	20
	Number of days of hospital capacity exceeded		0	0	12	24	41
Scenario 3 & 5	Observed Prevalence (%) Actual Prevalence (%)		3.299 5.150	5.525 8.535	9.079 13.792	14.487 21.456	32.049 43.618

Continue to next page

Delays			95 days	100 days	105 days	110 days	120 days
Scenario 3	Number shutdown	0%	343	343	331	331	375
		25%	0	0	0	36	0
		50%	0	0	45	0	0
		75%	57	57	24	33	25
	Number of days of hospital capacity exceeded		0	0	14	26	41
Scenario 5	Number shutdown	0%	305	295	307	343	374
		25%	24	49	44	0	6
		50%	71	41	33	38	0
		75%	0	15	16	19	20
	Number of days of hospital capacity exceeded		0	0	13	25	40

S.3. Summary of Optimal Policy for Scenarios 6 and 7

Table S2: Summary of scenarios 6 and 7. Shaded cells are the optimal policy in terms of when to be implemented (number of days of delays) and how to be implemented (number of days in each lockdown categories).

Delay			30 days	45 days	60 days	75 days	90 days
Scenario 6	Observed Prevalence (%) Actual Prevalence (%)		0.002 0.0032	0.01 0.016	0.054 0.085	0.272 0.43	1.359 2.134
	Number shutdown:	0%	118	118	120	75	90
		25%	0	0	0	74	59
		50%	0	0	0	0	0
		75%	282	282	280	251	251
	Number of days of hospital capacity exceeded		0	0	0	0	0
	Number of deaths		152	152	159	264	789

Continue to next page

Delay			**30 days**	**45 days**	**60 days**	**75 days**	**90 days**
Scenario 7	Observed Prevalence (%) Actual Prevalence (%)		0.0025 0.004	0.0139 0.0223	0.0741 0.118	0.384 0.610	1.946 3.057
	Number shutdown:	0%	54	53	67	75	143
		25%	47	52	53	66	0
		50%	0	0	0	0	0
		75%	299	295	280	259	257
	Number of days of hospital capacity exceeded		0	0	0	0	0
	Number of deaths		1935	1936	1959	2077	2703
Delays			**95 days**	**100 days**	**105 days**	**110 days**	**120 days**
Scenario 6	Observed (%) Actual (%)		2.305 3.60	3.878 6.019	6.438 9.877	10.465 15.763	24.787 34.90
	Number shutdown	0%	95	150	105	110	172
		25%	50	0	65	90	0
		50%	0	0	0	0	0
		75%	255	250	230	200	228
	Number of days of hospital capacity exceeded		0	0	0	16	35
	Number of deaths		1219	1992	2971	4374	7912

Continue to next page

Delays			**95 days**	**100 days**	**105 days**	**110 days**	**120 days**
Scenario 7	Observed Prevalence (%) Actual Prevalence (%)		3.299 5.15	5.525 8.535	9.079 13.792	14.487 21.456	30.86 42.29
	Number shutdown	0%	144	152	169	195	303
		25%	0	0	0	0	0
		50%	0	0	0	0	0
		75%	256	248	231	205	97
	Number of days of hospital capacity exceeded		0	0	14	26	40
	Number of deaths		3064	3755	4775	6136	9607

References

1 V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, *et al.*, Playing Atari with deep reinforcement learning, preprint, arXiv: 1312.5602.

AIMS Press