

SIMULATION-BASED REINFORCEMENT LEARNING FOR SOCIAL DISTANCING

by

F. Ege Hoşgüngör

213150

A Dissertation

Submitted to the Department of Engineering & Informatics

University of Sussex

In Partial Fulfilment of the Requirements

For the Degree of Master of Science

June 2020

TABLE OF CONTENTS

[1 ACKNOWLEDGEMENT 2](#_Toc47556012)

[2 ABSTRACT 3](#_Toc47556013)

[3 INTRODUCTION 3](#_Toc47556014)

[4 RELATED WORK 4](#_Toc47556015)

[Flocking 4](#_Toc47556016)

[5 METHODOLOGY 5](#_Toc47556017)

[6 SOCIAL DISTANCING 5](#_Toc47556018)

[7 THEORETICAL FRAMEWORK 5](#_Toc47556019)

[8 RESEARCH DESIGN 5](#_Toc47556020)

[9 POLICY OPTIMIZATION 5](#_Toc47556021)

[10 RESULTS 5](#_Toc47556022)

[11 DISCUSSION AND CONCLUSION 5](#_Toc47556023)

[12 REFERENCES 5](#_Toc47556024)

[13 Appendix 5](#_Toc47556025)

# ACKNOWLEDGEMENT

The report may be freely copied and distributed provided the source is acknowledged.

Throughout the writing of this dissertation I have received a great deal of support and assistance. I would first like to thank my supervisor, Prof. T. Nowotny, whose expertise was invaluable in the formulating of the research topic and methodology in particular.

In addition, I would like to thank my parents for their wise counsel and hospitality during these rough times. You are always there for me. Finally, there are my friends, who were of great support in deliberating over our problems and findings, as well as providing happy distraction to rest my mind outside of my research.

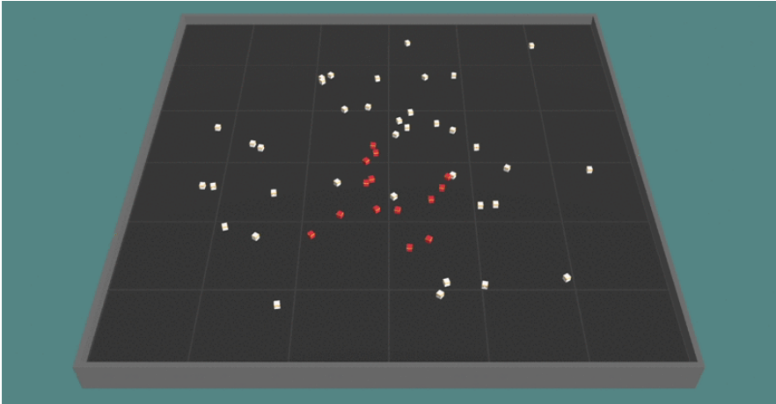
# ABSTRACT

Through agent-based modelling, and standard reinforcement learning algorithms at scale, we found AI agents can give insights about ongoing epidemics by simulating the disease. An epidemic simulation has been created with a physics engine and analyzed its results with SIR graphs. We found clear evidence of the relation between social distancing and getting infected rates. We further provide evidence that multi agent cooperation may scale better with increasing environment complexity and lead to a behavior that closer to far more human behavior. Ya da maskenin takılma yüzdesi ne kadar etkiliyor bunu yazıcaz.

<https://www.scribbr.co.uk/thesis-dissertation/abstract/>

# INTRODUCTION

Ever since the outbreak of Severe Acute Respiratory Syndrome as known as Covid-19 came out, life has changed drastically. Many researchers dedicated themselves to fight against the spread of this fatal virus and minimize the loss. Artificial Intelligence researchers are focusing their expertise knowledge to develop mathematical models for analyzing this epidemic disease.[1] An epidemic disease requires quick decisions to be made about interventions that could reduce or contain the disease spread. Decision makers need to be agile about their strategy since they race with the time. Every second that is wasted doubles the damage to humanity. In contrast, in order to decide confidently which strategy will work, decision makers need to analyze many scenarios and variables since if they give a wrong decision, that can also cause harm. This is an optimization problem which researchers cope by creating their own data and utilizing Reinforcement Learning to develop optimal strategies.[2]

Reinforcement Learning is an area of Machine Learning where an agent learns the best behavior by interacting with the environment. Creating these complex environments and artificial intelligent agents that solve complex human relevant tasks has been life-long challenge for RL researchers.[3] Environment which is a crucial component of RL, describes the task that the agent attempts to solve. Agent and environments cannot be considered separately, and it is only a design choice for the researcher to determine where the environment starts and the agent ends. Most of the time environment is defined as anything that agent cannot have a direct control on it. In this article we introduce an environment which can simulate the epidemic spread with physics engine of Unity. With the help of Reinforcement Learning, agents are trained to learn social distancing by their own.

*Figure 1:The Environment and agents. The square shape court limits the area and the cubes are the agents. White ones represent healthy and red ones represent infectious agents. The simulation shows how one infected agent starts to spread the disease in each time step.*

Social distancing is one of the most effective precaution that individuals can apply to their daily life. It is a successful strategy to prevent the infectious disease from spreading. It has many forms but at its core, the aim is to keep people apart enough from each other by putting physical distance and/or confining them to their homes. In this article, we took the idea of social distancing and simplified it to physical interaction. In other words, we assume social distancing is just having a physical distance between individuals. An infection mechanism starts when individuals are closer to each other than the threshold value. Being closer means individual has higher chance to get infected. In following chapter, this mechanism will be explained in detail.

To analyze possible outcomes, we used SIR graphs which is a widely common mathematical model that provides insights about the infectious disease outbreak. The model divides individuals to three categories. Susceptible, Infectious and Recovered. Recovered ones can be also called as Removed since they don’t have any affect to simulation. We also used the same number of categories in our simulation. Although it has been proven that the individual has short-term immunity after recovering from the covid-19, the character of the virus is still ambiguous. To avoid ambiguity, we assume when individuals recovered from the disease, they will immune to it forever and after recovery, they don’t have any effect to the curve.

For our deep reinforcement learning task, we used Proximal Policy Optimization Algorithm (PPO). This algorithm is designed by OpenAI. [3] and it is used in numerous different tasks from robotics to atari games. On a collection of benchmark tasks, PPO outperformed other online policy gradient methods and had a better balance between sample complexity and simplicity. OpenAI defines their algorithm with three features. Easy code, sample efficiency and ease of tune.

We introduce a new cooperative multi-agent physics-based reinforcement learning environment for control of epidemic spread. Through only a health status-based reward function, agents learn many human-relevant skills to protect themselves from epidemic outbreak including social distancing and self-isolation when they get sick. For example, agents learned how to maintain a balance between collecting reward boxes and not risking getting infected. We find that for several tests, curriculum learning can visibly change the results of the training. Setting the task step by step harder helped agents to learn better and to converge the loss function closer to the global minimum. In addition to that we showed pretrained agents learned the task faster than agents which is trained from scratch. Moreover, we observe signs of collaboration and simple communication between agents even though they don’t get direct reward from their actions. For instance, in one of the training, infected agents learned to gather up in a location where they avoid infecting others without knowing their health status.

The main contributions of this work are: 1) clear evidence that social distancing is mathematically a correct way to flatten the SIR curve. 2) A demonstration of how agent-based strategies and advances in computing can be leveraged to determine the optimal policy in an epidemic outbreak for a particular environment without expert human guidance.[4] 3) Evidence that self-isolation buralara bir şeyler gelmeli ve acil gelmeli



*Figure 2: Different states of the agents. Through the simulation, agent’s status of health changes. To represent the change, we used 4 different colors. a) White Bots indicates that the agent is not controlled by a brain. It only has simple hard-coded actions such as directly going targeted locations or bouncing from the walls. This represent individuals in a community who are not acting logically. b) Blue Bots are agents with a brain which controls them. c)Red Bots indicates*

*whether with brain or not the bot is infected. d)Purple indicates that agent is not infectious anymore. In SIR models’ purple agents call as Recovered-Removed. Please see* [[*https://github.com/Hsgngr/Pandemic\_Simulation*](https://github.com/Hsgngr/Pandemic_Simulation)](https://openai.com/blog/emergent-tool-use)*for example videos.*

# RELATED WORK

The history of agent-based modelling can be traced back to the The Simula programming language, which is developed in the mid-1960s and widely used as the first framework for automating step-by-step agent simulations.[5] Agent-based modelling use simple rules which can result in different sort of complex behavior. These models consist of interacting rule-based agents to create real-world-like complexity. The problem is the rules are strictly defined by researchers as hard-coded and therefore it is hard to generalize the simulation and get interesting behaviors.

Early work explored

Was further explored

More recent work attempted

In the context of Multi agent

Multi agent (cooperation) -Food Collection

Pandemic Simulation -SIR Model RL in Covid

Flocking- gibi başka distance ayarlama algoritmaları

# Flocking

# METHODOLOGY

* 1. SOCIAL DISTANCING

Social distancing’in mekanizmasını anlat. Proximity mechanism

* 1. THEORETICAL FRAMEWORK

The theoretical framework is the structure that can hold or support a theory of a research study. The theoretical framework introduces and describes the theory which explains why the research problem under study exists.

* 1. RESEARCH DESIGN

Environment for Epidemic Control

* 1. POLICY OPTIMIZATION

Agent policies are optimized using Proximal Policy Optimization (PPO) and Soft-Actor Critic (SAC). Both algorithms are compared, and PPO is selected for this task. The training is performed using Unity engine and open source Unity ML-Agents Toolkit. The agents trained in single-agent environments which is located next to each other in scene but do not have any interaction between them. At execution time, each agent act by using only their own observations and at optimization time, we use all agents’ observations to update our policy. So even though 8 different environments are used during training there was only neural network as an output. In other words, agents share the same policy parameters but act and observe independently as each of them were in different states.

# RESULTS

# DISCUSSION AND CONCLUSION

# REFERENCES

[1] N. S. Punn, S. K. Sonbhadra, and S. Agarwal, “COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms,” *medRxiv*, p. 2020.04.08.20057679, 2020.

[2] A. Yañez, C. Hayes, and F. Glavin, “Towards the control of epidemic spread: Designing reinforcement learning environments,” *CEUR Workshop Proc.*, vol. 2563, pp. 188–199, 2019.

[3] B. Baker *et al.*, “Emergent Tool Use From Multi-Agent Autocurricula,” 2019.

[4] O. Bent, S. L. Remy, S. Roberts, and A. Walcott-Bryant, “Novel exploration techniques (NETs) for malaria policy interventions,” *32nd AAAI Conf. Artif. Intell. AAAI 2018*, pp. 7735–7740, 2018.

[5] B. C. Perley, “Programming Language Maintenance,” *Defying Maliseet Lang. Death*, pp. 63–84, 2017.

# Appendix