Democratic Republic of Algeria and Working-Class

Ministry of High Education and Scientific Research

University Saad Dahlab Blida 1

Faculty of Science

Computer Department

PAPER FOR OBTAINING THE MASTER 2 DIPLOMA IN COMPUTER SCIENCE

Option: Network Computer Systems

Theme

Design and production of a tool for estimating and evaluating the risk of contamination by the COVID-19 virus in closed areas

Realized by: DIF MOHAMED NADIR

Tutor: Yassine Douga

College year 2021/2022

Remerciements

Nous commencerons par remercier nos familles et amis pour nous avoir soutenu et encouragé durant la réalisation de ce projet.

Nous adressons notre reconnaissance tout particulièrement à Mr Yacine Douga pour avoir été notre promoteur de projet.

Nous tenons à remercier les membres du jury pour avoir pris le temps d’examiner et évaluer notre travail.

Nous n’oublierons pas les professeurs qui nous ont suivi tout au long de notre parcours universitaire, nous leur sommes reconnaissants pour leurs efforts et le savoir qu’ils nous ont transmis.

Contents

Shape table

Acronym table

Symbol	Meaning
Q	Exhaled air flow
f	Fraction that crosses the mask
V	Room volume
λ, Lambda	Air Renewal rate
t, tau	Exposure time
N	Number of persons
P	Probability of virus infection
C_v	Virus concentration in exhaled air
C_q	Quantum concentration of infection in exhaled air
RL	Reinforcement Learning
S1 S2 S3	The past three weeks transmission estimation
rp	Current week transmission estimation
pc	Total transmission estimation
tp	Total probability of transmission
COVID-19	Corona virus 2019
SARSA	State-Action-reward- State-Action
DQN	Deep Q Network
DDPG	Deep Deterministic Policy Gradient
TRPO	Trust Region Policy Optimization
PPO	Proximal Policy Optimization
QR-DQN	Distributional Reinforcement Learning With Quintile Regression
HER	Hindsight Experience Replay
MDP	Markov Design Process
DP	Design Programming
TD	Temporal Difference Algorithms

Tables

Abstract

Since the end of 2019, the COVID19 virus has appeared in the world. This virus has left and still leaves behind thousands of deaths. Several security measures have been adopted to limit the spread of the latter among the world's populations. Following scientific research, in April 2021, American researchers proposed a formula that estimates and limits the risk of contagion by COVID19. One of the peculiarities of COVID19 is that it is a virus that continues to evolve and mutate. A theme aimed at developing an application based on the formula proposed by [1] to design and production of a tool for estimating and evaluating the risk of contamination by the COVID-19 virus in closed areas.

Keywords: COVID-19, Machine Learning, Artificial Intelligence, Reinforcement Learning.

Résumé

Depuis la fin de l’année 2019, le virus COVID-19 a fait son apparition dans le monde. Ce virus a laissé et laisse encore derrière lui des milliers de morts. Plusieurs mesures de sécurité ont été adoptées afin de limiter la propagation de ce dernier au sein des populations mondiales. Suite à une recherche scientifique, en Avril 2021, des chercheurs américains ont proposé une formule qui permet d’estimer et de limiter le risque de contagion par le COVID19. Une des particularités du COVID19 est que c’est un virus qui ne cesse d’évoluer et de muter. Un thème visant à développer une application qui se base sur la formule proposée par [1] pour concevoir et réaliser un outil d'estimation et d'évaluation du risque de contamination par le virus COVID-19 en milieu fermé.

Mots-clés :COVID-19, machine Learning, Intelligence Artificielle, Apprentissage par Renforcement.

ملخص

منذ نهاية عام 2019، ظهر فيروس COVID19 في العالم. لقد خلف هذا الفيروس ولا يزال وراءه آلاف الوفيات. وقد تم اتخاذ العديد من الإجراءات الأمنية للحد من انتشار هذا الأخير بين سكان العالم. بعد البحث العلمي ، في أبريل 2021 ، اقترح باحثون أمريكيون صيغة تقدر وتحد من خطر العدوى بواسطة COVID19. تتمثل إحدى خصائص COVID19 في أنه فيروس يستمر في التطور والتحول.فإن الموضوع الذي يهدف إلى تطوير تطبيق بناءً على الصيغة المقترحة بواسطة [1] لتصميم وإنتاج أداة لتقدير وتقييم مخاطر التلوث بفيروس COVID-19 في المناطق المغلقة.

الكلماتالمفتاحية:كوفيد -19 ، التعلم المعزز ،التعلمالآلي،الذكاءالاصطناعي

GENERAL INTRODUCTION

The COVID19 virus has appeared in the world this virus has left and still leaves behind thousands of deaths one of the peculiarities of COVID19 is that it is a virus that continues to evolve and mutate spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes. [32]

The problematic is that the Six-Foot Rule, a guideline that offers little protection from pathogen-bearing aerosol droplets sothe importance of airborne transmission of COVID-19 is now widely recognized. While tools for risk assessment have recently been developed, no safety guideline has been proposed to protect against it. Which is known to be transported by respiratory droplets exhaled by an infected person other problematic isThe Six-Foot Rule, a guideline that offers little protection from pathogen-bearing aerosol droplets, No safety guideline has been proposed to protect against airborne transmission of COVID-19, also there is another problematic which is the difficulty of Prediction and estimation of the risk of contamination by the COVID19 virus in closed areas. [1]

In this thesis, we will propose an original approach to design and produce a tool for estimating and evaluating the risk of contamination by the COVID19 virus in closed areas.

In **Chapter 1**, we present an introduction to the terms “Artificial Intelligence” and a brief definition of COVID-19.

In **Chapter 2**, we explain the concept of reinforcement learning and some basics and the downsides of reinforcement learning. Next, we take a tour of Reinforcement Learning-based algorithms, finally a comparison of the discussed algorithms.

In **Chapter 3**, we present the proposed solution the conception and our contribution to the proposed solution and the integration of Reinforcement Learning in the solution, and the steps taken to realize all this.

In **Chapter 4**, we give the results obtained from the tests and the scenarios tested.

Chapter 1 General Concepts

1. Introduction

Concurrent advances in information technology infrastructure and mobile computing have raised hopes that artificial intelligence (AI) might help to address challenges unique to the field of global health and accelerate achievement of the health-related sustainable development goals. AI-driven health interventions fit into four categories: diagnosis, patient morbidity or mortality risk assessment, disease outbreak prediction and surveillance, and health policy and planning. [33]

2. Covid-19

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus.

Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness.

The best way to prevent and slow down transmission is to be well informed about the COVID-19 virus, the disease it causes and how it spreads.

There are thought to be three possible routes of human-to-human transmission of COVID-19: large drop transmission from the mouth of an infected person to the mouth, nose or eyes of the recipient; physical contact with droplets deposited on surfaces (fomites) and subsequent transfer to the recipient's respiratory mucosae; and inhalation of the micro droplets ejected by an infected person and held aloft by ambient air currents [9–10]. They subsequently refer to these three modes of transmission as, respectively, "large-drop," "contact," and "airborne" transmission, while noting that the distinction between large-drop and airborne transmission is somewhat nebulous given the continuum of sizes of emitted droplets [11]. They build on models of airborne disease transmission in order to derive an indoor safety guideline that would impose an upper bound on the "cumulative exposure time," the product of the number of occupants and their time in an enclosed space. theydemonstrate how this bound depends on the rates of ventilation and air filtration, dimensions of the room, breathing rate, respiratory activity and face mask use of its occupants, and infectiousness of the respiratory aerosols.

2.1 A guideline to limit indoor airborne transmission of COVID-19

The pathogen responsible for COVID-19, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is known to be transported by respiratory droplets exhaled by an infected person [12–13]. There is now overwhelming evidence that indoor airborne transmission associated with relatively small, micron-scale aerosol droplets plays a dominant role in the spread of COVID-19 [17, 18, 19, 20–21, 22], especially for so-called "super-spreading events" [23–24], which invariably occur indoors [25].

2.1.1 The Well-Mixed Room

They assume that the droplet-borne pathogen remains airborne for some time before being extracted by the room's ventilation system, inhaled, or sedimenting out. The fate of ejected droplets in a well-mixed ambient is determined by the relative magnitudes of two speeds: the settling speed of the drop in quiescent air and the ambient air circulation speed within the room, they consider a well-mixed room of area A, depth H, and volume V =HA with ventilation outflow rate Q and outdoor air change rate (typically reported as air changes per hour, or ACH) _a =Q=V. Mechanical ventilation imposes an additional recirculation flow rate $\boldsymbol{C_r}$ that further contributes to the well-mixed state of the room, but alters the emergent drop size distributions only if accompanied by filtration. It is noteworthy that, even in the absence of forced ventilation, there will generally be some mixing in an enclosed space: Natural ventilation will lead to flows through windows and doors, as well as leakage through construction materials and joints. Moreover, occupants serve to enhance airflow through their motion and respiration. [1]

2.1.2 Indoor Safety Guideline

The concentration of infection quanta or "infectiousness" of exhaled air, $\boldsymbol{C_q}$the latter is the key disease-specific parameter in their model, which can also be expressed as the rate of quanta emission by an infected person, they thus arrive at a simple guideline, appropriate for steady-state situations, To minimize risk of infection, one should avoid spending extended periods in highly populated areas. One is safer in rooms with large volume and high ventilation rates. One is at greater risk in rooms where people are exerting themselves in such a way as to increase their respiration rate and pathogen output, for example, by exercising, singing, or shouting. Since the rate of inhalation of contagion depends on the volume flux of both the exhalation of the infected individual and the inhalation of the susceptible person, the risk of infection increases asQ^2. [1]

2.1.3 Application to COVID-19

They proceed by making rough estimates for $C_q$for different respiratory activities on the basis of existing epidemiological data gathered from early super-spreading events of COVID-19. Their inferences provide a baseline value forC_q, relevant for elderly individuals exposed to the original strain of SARS-CoV-2, an inference of C_q=970 quanta/m3 was made by Miller et al. [23] in their recent analysis of the Skagit Valley Chorale super-spreading incident [26], on the basis of the assumption that the transmission was described in terms of the Wells–Riley model [27, 28, 20, 29] for a mean breathing rate. This inference is roughly consistent with studies of other related viral diseases. For example, Liao et al. [30] estimated C_q=28 quanta/m^3 from the rate of indoor spreading of SARS-CoV-2, in a hospital and an elementary school. Estimates of $C_q$for H1N1 influenza fall in the range 15 to 128 quanta/m^3 [31]. For SARS-CoV-2, Buonanno et al. [20] estimate a $C_q$range of 10.5 to 1,030 quanta/m^3, on the basis of the estimated infectivity and note that the precise value depends strongly on the infected person’s respiratory activity, as it shown in Figure 13 below.

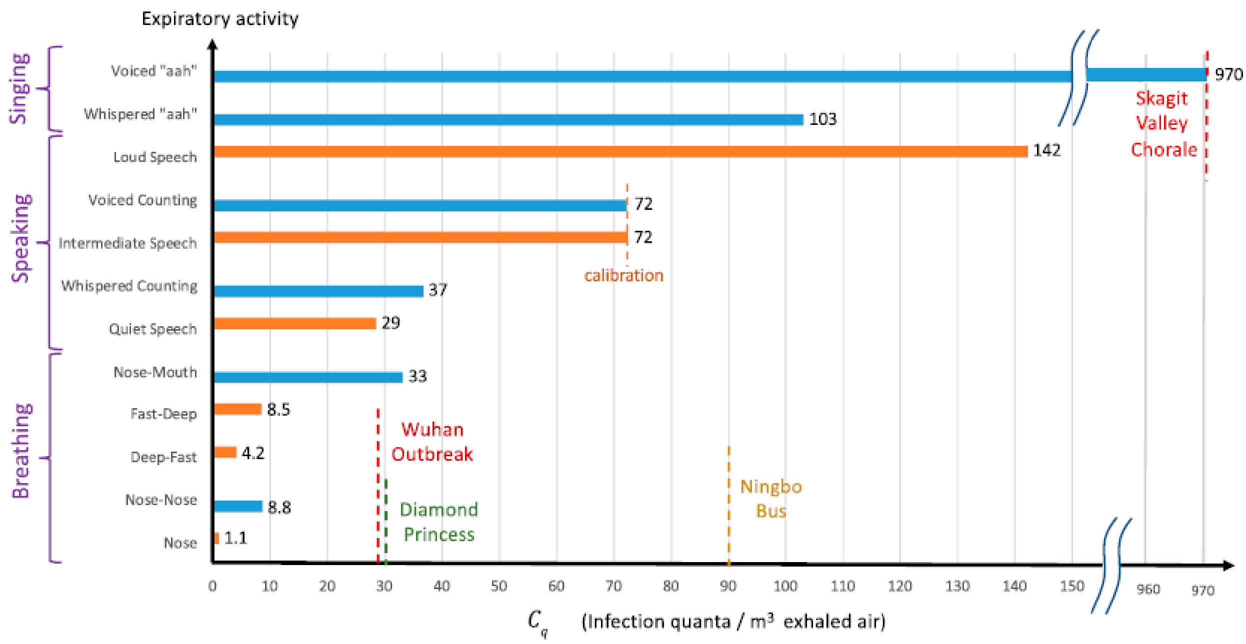


Figure 13: Estimates of the "infectiousness" of exhaled air, C_q, defined as the peak concentration of COVID-19 infection quanta in the breath of an infected person, for various respiratory activities, obtained from the experiments on hundreds of people .[1]

2.1.4Discussion and Caveats

They have focused here primarily on airborne transmission, for which infection arises through inhalation of a critical quantity of airborne pathogen, and neglected the roles of both contact and large-drop transmission [14]. While motivated by the COVID-19 pandemic, their theoretical framework applies quite generally to airborne respiratory illnesses, including influenza. Moreover, they note that the approach taken, coupling the droplet dynamics to the transmission dynamics, allows for a more complete description. For example, consideration of conservation of pathogen allows one to calculate the rate of pathogen sedimentation and associated surface contamination, consideration of which would allow for quantitative models of contact transmission and so inform cleaning protocols. Respiration rates Q have been measured to be 0.5 m^3/h for normal breathing, and may increase by a factor of 3 for more strenuous activities [20]. Other parameters, including room geometry, ventilation, and adherence to the Six-Foot Rule would limit large-drop transmission, and adherence to their guideline. Above all, their study makes clear the inadequacy of the Six-Foot Rule in mitigating indoor airborne disease transmission, and offers a rational, physically informed alternative for managing life in the time of COVID-19. If implemented, their safety guideline would impose a limit on the CET in indoor settings, violation of which constitutes an exposure for all of the room’s occupants. Finally, while their study has allowed for an estimate of the infectiousness of COVID-19, it also indicates how new data characterizing indoor spreading events may lead to improved estimates thereof and so to quantitative refinements of their safety guideline. [1]

2.1.5 Modes of Contamination

First theymentioned the transmission of COVID by air a form of transmission in which you will see that it is largely based on physical principles and the project objective it's to see at best how to limit this mode of transmission, at the present hour they mainly distinguish 3 modes of transmission of Sars-Cov-2, the virus responsible for COVID, first of all there is the mode that they could call "large droplets" the idea is known: when we cough, sneeze or spit we will project droplets that contain the virus, if these droplets land in the mouth, nose or eyes from another person, we are likely to infect them, they know that these droplets have sizes that can reach a millimeter and as they are relatively big they will fall under the effect of gravity they therefore estimate on the basis of various visualizations that these droplets wind to be projected at most about 2m, this is the reason why they generally recommend this minimum distance As a barrier gesture and of course also an important motivation for wearing the mask which blocks the jet of droplets. Another possible mode of contamination: via surfaces, if droplets land for example on a table that I then come to touch, I will bring back the virus on my hands and if I put them in my mouth, nose or eyes, I will contaminate myself, the barrier gesture for this mode of contamination as we know it: washing your hands regularly with soap, using hydro-alcoholic gel it doesn't get any better and disinfect surfaces likely to be contaminated. This mode of transmission by surfaces, which had been envisaged very early, they think today that it is certainly not a very active mode that there is a lack of direct evidence that it acts in a truly meaningful way. On the other hand, with the evolution of their understanding of the mechanisms it is now believed that a major or even dominant mode of transmission it is transmission by air, by aerosol, they talked about the big droplets, but it's not just that, when we cough, but also when we speak and even just our breathing we emit a whole set of much smaller droplets the size of which can range from a hundred microns to less than one micron these droplets come from our saliva and the mucous membranes of the lungs and respiratory tract and even if you don't have symptoms like rate or sneezing. You will naturally emit these droplets. However, unlike the large droplets they talked about just before these can, if they are small enough, they can stay suspended in the air. The ambient air movements are sufficient to compensate for their fall under the effect of gravity. For these micro-droplets suspended in the air, the term aerosols is used, and these suspended aerosols can diffuse freely in a closed room, and contaminate a person, even if he is more than two meters from the carrier of the virus. Obviously, there is no strict border between the droplets they talked about and the aerosols. The larger the droplets, the faster they will fall to the ground, the smaller they are, the more likely they are to stay in suspension. In a closed room, the diffusion of contaminated aerosols to the whole room can be quite fast. The natural movements of the air, the movements of people, make that after a few minutes at best, they can almost consider the aerosol concentration to be homogeneous all over the room. This would mean that for this mode of contamination, the distance does not matter. Whether you are close enough to the carrier or across the room, the risk will be similar. [1]

3. Artificial intelligence (AI)

Is a technical science that studies and develops theories, methods, technologies, and applications for simulating and extending human intelligence. The purpose of AI is to enable machines to think like people and to make machines intelligent. Today, AI has become an interdisciplinary course that involves various fields, as it shown in Figure 1 below.[2]

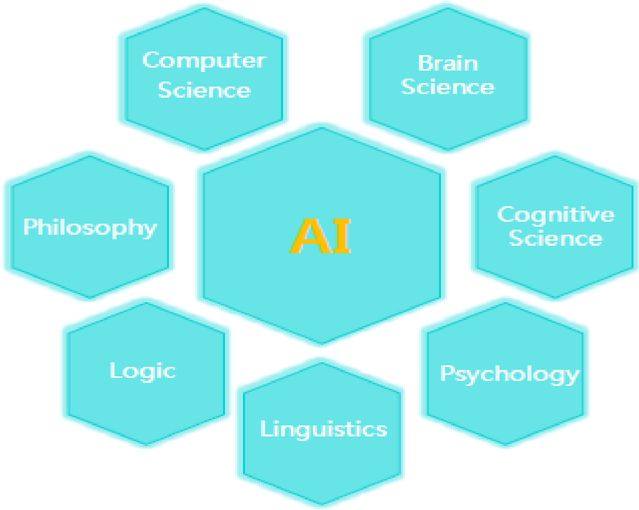


Figure1: AI [2]

Artificial intelligence (AI) is being used as a tool to support the fight against the viral pandemic that has affected the entire world since the beginning of 2020. The press and the scientific community are echoing the high hopes that data science and AI can be used to confront the coronavirus. [34]

China, the first epicenter of this disease and renowned for its technological advance in this field, has tried to use this to its real advantage. Its uses seem to have included support for measures restricting the movement of populations, forecasting the evolution of disease outbreaks and research for the development of a vaccine or treatment.

With regard to the latter aspect, AI has been used to speed up genome sequencing, make faster diagnoses, carry out scanner analyses or, more occasionally, handle maintenance and delivery robots. [34]
Its contributions, does not eliminate the need for clinical test phases nor does it replace human expertise entirely. [34]

3.1 The contribution of artificial intelligence to the search for a cure

The first application of AI expected in the face of a health crisis is certainly the assistance to researchers to find a vaccine able to protect caregivers and contain the pandemic. Biomedicine and research rely on a large number of techniques. [34]
The predictions of the virus structure generated by AI have already saved scientists months of experimentation. AI seems to have provided significant support in this sense. The American start-up Moderna has managed to significantly reduce the time required to develop a prototype vaccine testable on humans thanks to the support of bioinformatics, of which AI is an integral part. [34]
Similarly, Chinese technology giant Baidu, in partnership with Oregon State University and the University of Rochester, published its Linear-fold prediction algorithm in February 2020 to study the same protein folding. This algorithm is much faster than traditional algorithms in predicting the structure of a virus secondary ribonucleic acid (RNA) and provides scientists with additional information on how viruses spread. Deep-Mind, a subsidiary of Google's parent company, Alphabet, has also shared its predictions of coronavirus protein structures with its Alpha-Fold AI system. IBM, Amazon, Google and Microsoft have also provided the computing power of their servers to the US authorities to process very large datasets in epidemiology, bioinformatics and molecular modeling. [34]

3.2 Artificial intelligence, a driving force for knowledge sharing

Indeed, in the weeks following the appearance of the new coronavirus in Wuhan, China, in December 2019, nearly 2,000 research papers were published on the effects of this new virus, on possible treatments, and on the dynamics of the pandemic. [34]

Microsoft Research, the National Library of Medicine and the Allen Institute for AI (AI2) therefore presented their work on 16 March 2020, which consisted of collecting and preparing more than 29,000 documents relating to the new virus and the broader family of coronaviruses, 13,000 of which were processed so that computers could read the underlying data. [34]

3.3 Artificial intelligence, observer and predictor of the evolution of the pandemic

The Canadian company BlueDot is credited with the early detection of the virus using an AI and its ability to continuously review over 100 data sets. BlueDot detected what was then considered an outbreak of pneumonia in Wuhan, China on 31 December 2019 and identified the cities most likely to experience this outbreak.

A team of researchers working with the Boston Children's Hospital has also developed an AI to track the spread of the coronavirus. Called Health-Map, the system integrates data from Google searches, social media and blogs, as well as discussion forums.[34]

The International Research Centre for Artificial Intelligence (IRCAI) in Slovenia, under the auspices of UNESCO, has launched an "intelligent" media watch on coronavirus called Corona Virus Media Watch which provides updates on global and national news based on a selection of media with open online information. [34]

3.4 Artificial intelligence to assist healthcare personnel

For their part, two Chinese companies have developed AI-based coronavirus diagnostic software. The Beijing-based start-up Infer-vision has trained its software to detect lung problems using computed tomography (CT) scans. Originally used to diagnose lung cancer, the software can also detect pneumonia associated with respiratory diseases such as coronavirus. At least 34 Chinese hospitals are reported to have used this technology to help them screen 32,000 suspected cases. [34]

The Alibaba DAMO Academy, the research arm of the Chinese company Alibaba, has also trained an AI system to recognize coronaviruses with an accuracy claimed to be 96%. According to the company, the system could process the 300 to 400 scans needed to diagnose a coronavirus in 20 to 30 seconds, whereas the same operation would usually take an experienced doctor 10 to 15 minutes. The system is said to have helped at least 26 Chinese hospitals to review more than 30,000 cases. [34]

In South Korea, AI is reported to have helped reduce the time needed to design testing kits based on the genetic make-up of the virus to a few weeks, when it would normally take two to three months. The biotech company Seegene used its automated test development system to develop the test kit and distribute it widely. This has equipped 118 medical establishments with this device and tested more than 230,000 people. [34]

3.5 Artificial intelligence: an evaluation of its use in the aftermath of a crisis

Digital technology, including information technology and AI, are therefore proving to be important tools to help build a coordinated response to this pandemic. The multiple uses also illustrate the limits of what can currently be achieved by this very technology. [34]

4. Conclusion

In this chapter we discussed the terms of AI, COVID 19, different modes of contamination the facters helping the super spreading of this virus and the problems faced to protect against these modes of contamination so the conclusion here is that we must start paying more attention to the airborne transmission which is the number one factor to the super spreading events due to the less knowledge of this dangerous mode of contamination.

Chapter 2 Reinforcement Learning

1. Introduction

People from different backgrounds have started to wonder how to learn intelligent behavior in complex dynamic environments, so for this chapter we are going to answer this question by diving into the term of Reinforcement Learning by covering terminologies algorithms and the difficulties of the ladder.

2. Reinforcement Learning

Reinforcement learning it’s like teaching your dog to do tricks:if your pet performs the trick you desirethere will a reward, otherwise a penalty. [3-4]

RL deals with learning via interaction and feedback, learning to solve a task by trial and error, or in other words acting in an environment and receiving rewards for it. Essentially an agent (or several) is built that can perceive and interpret the environment in which is placed, furthermore, it can take actions and interact with it. [3-4]

3. Terminologies

As it shown in Figure 3 belowthe terminologies used in the field of RL are:

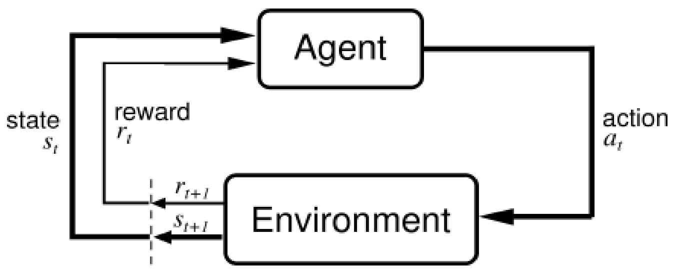


Figure 3:Agent-environmentinteractions [3]

Agent

The learner and the decision maker. [3]

Environment

Where the agent learns and decides what actions to perform. [3]

Action

A set of actions which the agent can perform. [3]

State

The state of the agent in the environment. [3]

Reward

For each action selected by the agent the environment provides a reward usually a scalar value. [3]

Policy

The decision-making function (control strategy) of the agent, which represents a mapping from situations to actions. [3]

4. RL in Depth

There are two significant downsides to this approach so on the one hand if you want to do supervised learning you have to create a data set to train on which is not always a very easy thing to do and on the other hand if you train your neural network model to simply imitate the actions of the human player well then by definition your agent can never be better at playing the game than that human gamer for example if you want to train a neural net to be better at playing the game of gold and the best human then by definition we can't use supervised learning so is there a way to have an agent learn to play a game entirely by itself well fortunately there is and this is called reinforcement learning so the framework and reinforcement learning is actually surprisingly similar to the normal framework in supervised learning so we still have an input frame we run it through some neural network model and the network produces an output action but the only difference here is that now we don't actually know the target label so we don't know in any situation because we don't have a data set to train on and in reinforcement learning the network that transforms input to output actions is called the policy network now one of the simplest ways to train a policy network is a method called policy gradients so the approach in policy gradients is that you start out with a completely random network you feed that network with an input it produces a random output action and the network in this case it could be a fully connected network but you can obviously apply convolution there as well and now in reality the output of your network is going to consist of two numbers and what will you do while training is actually sample from the distribution so that you're not always going to repeat the same exact actions and this will allow your agent to sort of explore the environment a bit randomly and hopefully discover better rewards and better behavior now importantly because we want to enable our agent to learn entirely by itself the only feedback that we're going to give it is the reward so whenever our agent manages to score a goal it will receive a reward of +1 and if he doesn't then our agent will receive the penalty of minus 1 and the entire goal of the agent is to optimize it's policy to receive as much reward as possible so in order to train our policy network the first thing is we're going to do is collect a bunch of experience so you're just going to run a whole bunch of those game frames through your network select random actions, as it shown in Figure 5 below

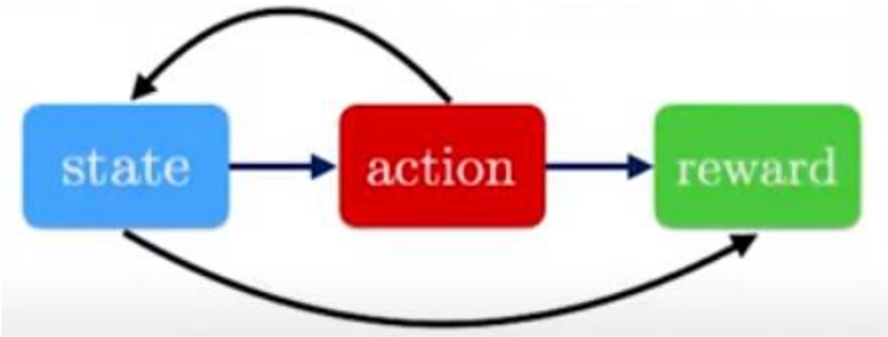


Figure 5: policy training[5]

and now obviously since our agent hasn't learned anything useful yet it's going to lose most of the games but the thing is sometimes our agent might get lucky sometimes it's going to randomly select a whole sequence of actions that actually lead to scoring a goal and in this case our agent is going to receive a reward and a key thing to understand that for every episode regardless of whether we want a positive or a negative reward we already compute the gradients that would make the actions that our agent has chosen more likely in the future and this is very crucial and so what policy gradients are going to do is that for every episode where we've got a positive reward we're going to use the normal gradients to increase the probability of those actions in the future but whenever we got a negative reward we're going to apply the same [5] gradient but we're going to multiply it with a minus and this minus sign will make sure that in the future all the actions that we took in a very bad episode are going to be less likely in the future, as it shown in Figure 6 below

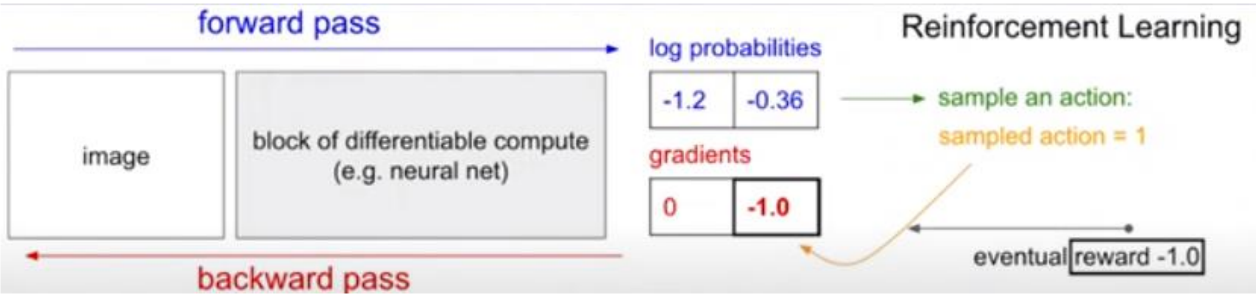


Figure 6: the rewards dynamiques[5]

And so the result is that while training our policy network the actions that lead to negative rewards are slowly going to be filtered out and the actions that leads to positive rewards are going to become more and more likely so in a sense our agent is learning. [5]

5. The significant downsides to using reinforcement learning

5.1 Credit Assignment Problem

The problem with policy gradients is that our policy gradients when it makes a mistake and gets a negative penalty so it's going to assume that since we lost that episode all of the actions that we took there must be bad actions and is going to reduce the likelihood of taking those actions in the future but the most part of that episode we were doing really well so we don't really want to decrease the likelihood of those actions [5] , as it shown in Figure 7 below

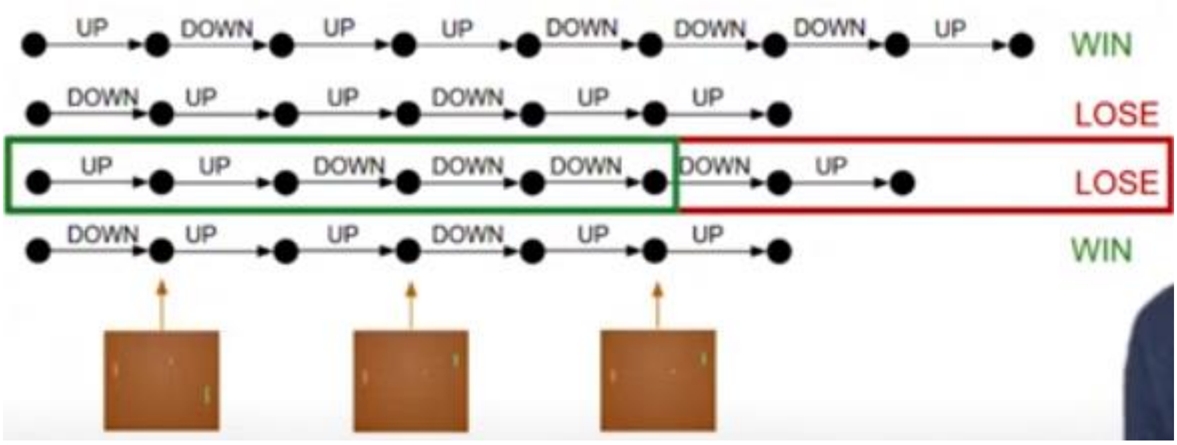


Figure 7: actions taken by agent in Ping-Pong game[5]

and in Reinforcement learning this is called the "Credit Assignment Problem" it's where if you get a reward at the end of your episode well what are the exact actions that led to that specific reward and this problem is entirely related to the fact that we have what we call a “Sparse Reward Setting”.[5]

5.2 Sparse Reward Setting

so instead of getting a reward for every single action we only get a reward after an entire episode and our agent needs to figure out what part of its actions sequence we're causing the reward that it eventually gets and so the result of this sparse reward setting is that in Reinforcement Learning algorithms are typically inefficient which means that you have to give them a ton of training time before they can learn some useful behavior now it turns out that in some extreme cases the sparse reward setting actually fails completely in the same case [5] whereas in reinforcement learning setting you're having to deal with this very big problem of sparse reward setting and so the traditional approach to solve this issue of sparse rewards has been the use of "Rewards Shaping".[5]

5.3 Rewards Shaping

so reward shaping is the process of manually designing a reward function that needs to guide your policy to some desired behavior by adding extra rewards to guide your policy to some desired behavior and while this makes it easier for your policy to converge to desired behavior there are some significant downsides to reward shaping so firstly reward shaping is a custom process that needs to be redone for every new environment you want to train a policy well you would have to craft a new reward function for every single environment that's just not scalable the second [5] problem is that reward shaping suffers from what we call "The Alignment Problem". [5]

5.4 The Alignment Problem

so it turns out that reward shaping is actually surprisingly difficult in a lot of cases when you shape your reward function your agent will find some very surprising way to make sure that it's getting a lot of rewards but not doing at all what you wanted to do and in sense the policy is just over fitting to that specific reward function that you designed while not generalizing to the intended behavior that you had in mind and there's a lot of funny cases where reward shaping goes terribly wrong so for example the agent was trained to do jumping and the reward function was the distance from its feet to the ground and what this agent has learned is to simply grow a very tall body and do some kind of backflip to make sure that its feet are very far from the ground. [5]

To give you one final idea of how hard it can be to the reward shaping in some cases like Alpha go for example by definition you don't want to do any reward shaping because this will constrain your policy the behavior of humans which is not exactly optimal in every situation.[5]

So the situation that we're now is that we know that it's really hard to train in a sparsely setting but at the same time it's also very tricky to shape a reward function and we don't always want to do that. [5]

6. Model-Free VSModel-Based Reinforcement Learning

Model-based RL uses experience to construct an internal model of the transitions and immediate outcomes in the environment. Appropriate actions are then chosen by searching or planning in this world model. The model stands for the simulation of the dynamics of the environment. That is, the model learns the transition probability $T(s1 | (s0, a))$ from the pair of current state $s0$ and action a to the next state s*1*. If the transition probability is successfully learned, the agent will know how likely to enter a specific state given current state and action. [6]

Model-free RLuses experience to learn directly one or both of two simpler quantities (state/ action values or policies) which can achieve the same optimal behavior but without estimation or use of a world model. Given a policy, a state has a value, defined in terms of the future utility that is expected to accrue starting from that state. [6]

Model-free methods are statistically less efficient than model-based methods, because information from the environment is combined with previous, and possibly erroneous.[6]

Model-free algorithms rely on trial-and-error to update its knowledge. As a result, it does not require space to store all the combination of states and actions. [6]

So **Model-based** learning attempts to model the environment then choose the optimal policy based on its learned model; In **Model-free** learning the agent relies on trial-and-error experience for setting up the optimal policy. [6]

7. On-policy vs Off-policy

An on-policy agent learns the value based on its current action a derived from the current policy, whereas its off-policy counterpart learns it based on the action a* obtained from another policy. In Q-learning, such policy is the greedy policy. [6]

8. Illustration of Various Algorithms

8.1 Q-Learning

Q-Learning is an off-policy, model-free RL algorithm,Q-learning learns the action-value function *Q(s, a)*: basically a scalar value is assigned over an action **a** given the state **S**. The following chart provides a good representation of the algorithm. [6]

As it shown in Figure 8 below

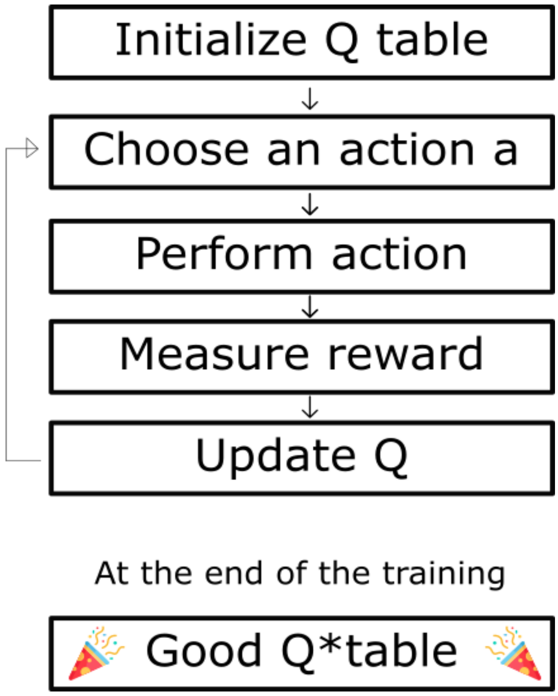


Figure8:Q-learning steps[6]

8.2 State-Action-Reward-State-Action (SARSA)

SARSA very much resembles Q-learning. The key difference between SARSA and Q-learning is that SARSA is an on-policy algorithm. It implies that SARSA learns the Q-value based on the action performed by the current policy instead of the greedy policy. [4]

8.3 Deep Q Network (DQN)

Although Q-learning is a very powerful algorithm, its main weakness is lack of generality. Q-learningresembles dynamic programming as updating numbers in a two-dimensional array (Action Space * State Space). This indicates that for states that the Q-learning agent has not seen before, it has no clue which action to take. [4]

Q-learning agent does not have the ability to estimate value for unseen states. [4]

To solve this problem, DQN get rid of the two-dimensional array by introducing Neural Network. [4]

DQN leverages a Neural Network to estimate the Q-value function. The input for the network is the current, while the output is the corresponding Q-value for each of the action, as it shown in Figure 9 below. [4]

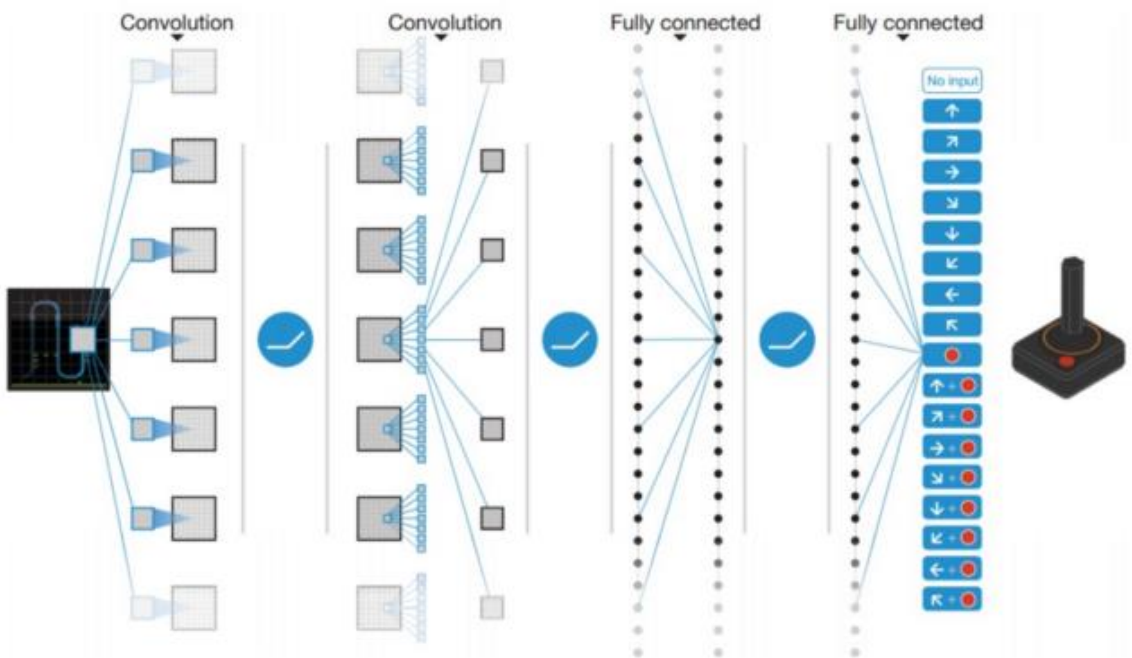


Figure9:Atari Example In 2013, Deep Mind applied DQN to Atari game, as illustrated in the above figure. The input is the raw image of the current game situation. It went through several layers including convolutional layer as well as fully connected layer. The output is the Q-value for each of the actions that the agent can take.[6]

8.4 Deep Deterministic Policy Gradient (DDPG)

Although DQN achieved huge success in higher dimensional problem, such as the Atari game, the action space is still discrete. However, many tasks of interest, especially physical control tasks, the action space is continuous. If you discretize the action space too finely, you wind up having an action space that is too large. For instance, assume the degree of free random system is 10. For each of the degree, you divide the space into 4 parts. You wind up having 4^{10} =1048576 actions. It is also extremely hard to converge for such a large action space. [4]

DDPG relies on the actor-critic architecture with two eponymous elements, actor and critic. An actor is used to tune the parameter $\boldsymbol{\theta}$ for the policy function, i.e. decide the best action for a specific state. [4]

A critic is used for evaluating the policy function estimated by the actor according to the temporal difference (TD) error, as it shown in Figure 10below. [4]

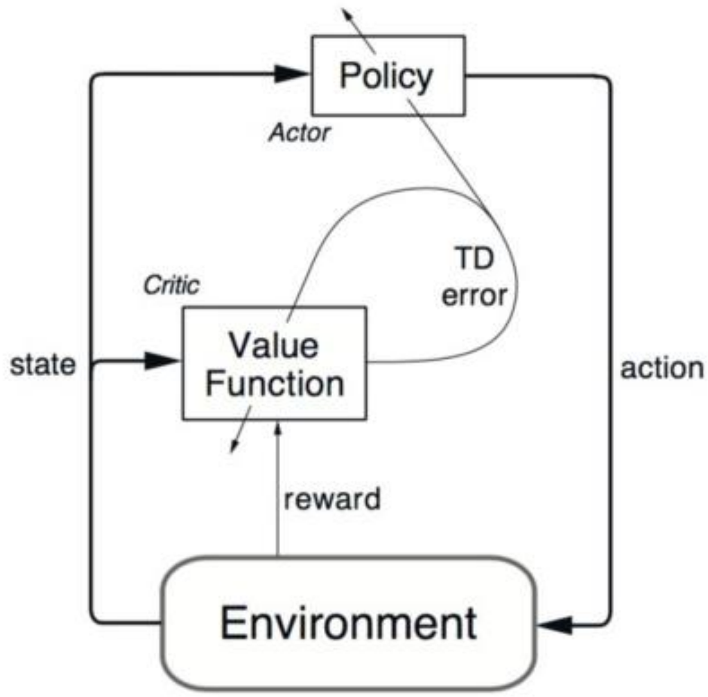


Figure10:Actor-critic Architecture[6]

8.5 Trust Region Policy Optimization (TRPO)

Deep Deterministic Policy Gradient (DDPG) was a break through that allows agent to perform actions in a continuous space while maintaining a descent performance. However, the main issue of DDPG is that you need to pick the step size that falls into the right range. If it is too small, the training progress will be extremely slow. If it is too large, conversely, it tends to be overwhelmed by the noise, leading to tragic performance. Recall that the target for calculating the Temporal Difference (TD) error is the following:

If the step size is selected inappropriately, the target *yi* derived from the networks or function estimators will not be good, leading to an even worse sample and worse estimate of the value function. [4]

Therefore, what we need is a way to update parameters that guarantees policy improvement. Namely, we want the **expected discounted long-term reward η** to be always increasing. [4]

8.6 Proximal Policy Optimization (PPO, OpenAI version)

Although TRPO has achieved great and consistent high performance, the computation and implementation of it is extremely complicated. In TRPO, the constraint imposed on the surrogate objective function is the KL divergence between the old and the new policy. [4]

Fisher Information Matrix, a second-order derivative of KL divergence, is used to approximate the KL term. This results in computing several second-order matrixes, which requires a great amount of computation. In the TRPO paper, Conjugate Gradient (CG) algorithm was used to solve the constrained optimization problem so that the Fisher Information Matrix does not need to be explicitly computed. Yet, CG makes implementation more complicated. [4]

PPO gets rid of the computation created by constrained optimization as it proposes a clipped surrogate objective function. [4]

rt($\boldsymbol{\theta}$) denotes the ratio between the new and the old policy. [4]

The idea of TRPO's constraint is disallowing the policy to change too much. [4]

- Case 1: When the advantage Ât is greater than 0

If Ât is greater than 0, it means that the action is better than the average of all the actions in that state. Therefore, the action should be encouraged by increasing rt ($\boldsymbol{\theta}$) so that this action has a higher chance to be adopted.

- Case 2: When the advantage Ât is smaller than 0

By contrast, if Ât is smaller than 0, then that action should be discouraged. As a result, rt ($\boldsymbol{\theta}$) should be decreased. As it shown in Figure 12 below.[4]

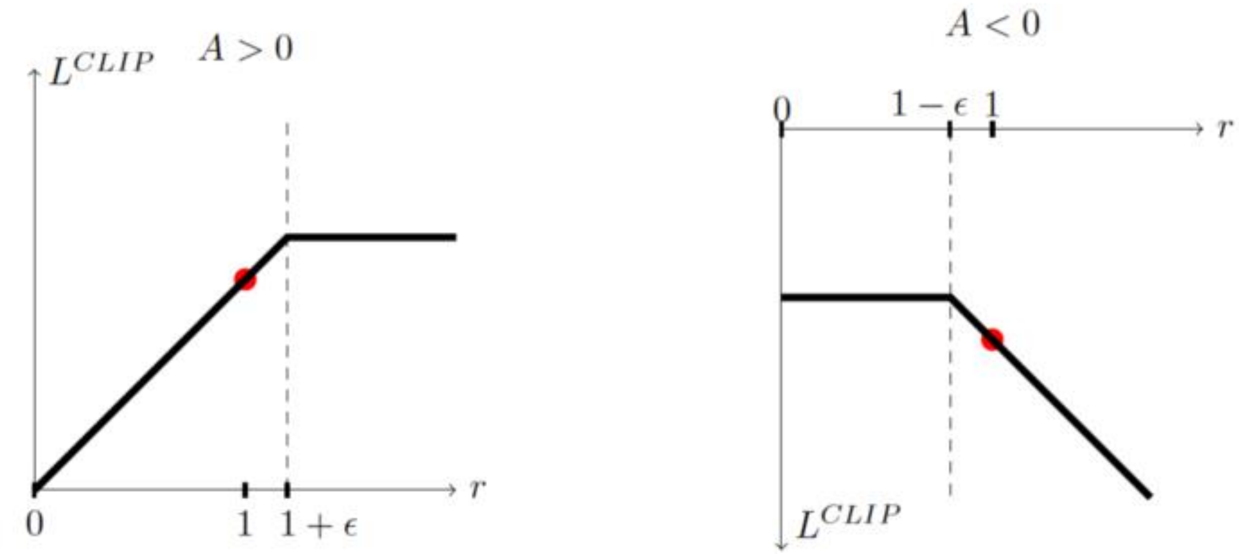


Figure12:Illustration of the Clip[6]

Essentially, it restricts the range that the new policy can vary from the old one; thus, removing the incentive for the probability ratio rt($\boldsymbol{\theta}$) to move outside the interval. [4]

In practice, loss function error and entropy bonus should also be considered during implementation. [4]

9. Model-based RL

Model-based RL has a strong influence from control theory, and the goal is to plan through an *f(s,a)* control function to choose the optimal actions. Think of it as the RL field where the laws of physics are provided by the creator. The drawback of model-based methods is that although they have more assumptions and approximations on a given task, but may be limited only to these specific types of tasks. There are two main approaches: learning the model or learn given the model.[6]

10. Comparison of Discussed Algorithms

Algorithm	Model	Policy	Action Space	Observation Space	Operator
Q-Learning	Model-Free	Off-Policy	Discrete	Discrete	Q-Value
SARSA	Model-Free	On-Policy	Discrete	Discrete	Q-Value
DQN	Model-Free	Off-Policy	Discrete	Continuous	Q-Value
DDPG	Model-Free	Off-Policy	Continuous	Continuous	Q-Value
TRPO	Model-Free	Off-Policy	Continuous	Continuous	Advantage
PPO	Model-Free	Off-Policy	Continuous	Continuous	Advantage

Table 1: Comparison Table[6]

All the discussed RL algorithms are model-free. That is, none of them are trying to estimate the objective function. Instead, they update their knowledge based on trial-and-error. Among all of them, only SARSA is on-policy, learning value based on its current action. DQN was a huge improvement from a discrete observation space to a continuous one, allowing the agent to handle unseen state. DDPG was another break through that enables agent to perform continuous actions with policy gradient, broadening the application of RL to more tasks such as control. TRPO improves the performance of DDPG as it introduces a surrogate objective function and a KL divergence constraint, guaranteeing non-decreasing long-term reward. PPO further optimizes TRPO by modifying the surrogate objective function, which improves the performance as well as decreasing the complexity of implementation and computation. [4]

The solution we chose is so much simpler than the ones mentioned above it requires a simple knowledge of math algebra and the basics of python, no data-set needed with a policy network so we don't need to go through the time consuming collecting the experience and training with sparse rewards which is really hard or the reward shaping which is also not an optimal solution by doing this we also got rid of the issues such as "Credit Assignment Problem" and "The Alignment Problem".

The main reason why not using any of the algorithms mentioned earlier is that they are basically used to train the agent how to play video games their functions procedures are specially made for video games even the data set contain a caption of video games movement made by human player.

11. Conclusion

This chapter we addressed “Reinforcement Learning” we discussed everything from definition all the different algorithms with the comparison of these discussed algorithms, the downsides the challenges facing RL so the reality is that most of these breakthroughs in "Reinforcement Learning" are actually the work of some of the brightest minds alive today and there is a lot of very hard engineering going behind the scenes.

Chapter 3 Conception and Implementation

1. Introduction

In the previous chapter we talked about RL in detail, in this one we are going how we integrated and modeled RL in the formula proposed by [1].

2. Proposed Solution

What [1] proposed was a formula for the risk of contamination by COVID-19 virus which is as follows$\frac{\mathbf{N}*\mathbf{t}*Q^2*f^2*\boldsymbol{C_q}}{\boldsymbol{\lambda}\mathbf{V}}$

First we have

Exhaled air flow: Qwhich has a double effect due to the inhaling and exhaling

Fraction that crosses the mask: **f**a well-worn mask only has 10% of the virus passing through each time which also has a double effect due to the inhaling and exhaling

Room volume: V

Air Renewal rate: **λ**or ventilation rate or in other words how many times the air is renewed in one hour

Exposure time: t

Number of persons: N

Quantum concentration of infection in exhaled air:$\boldsymbol{C_q}$the concentration of infection quanta or "infectiousness" of exhaled air, $\boldsymbol{C_q}$the latter is the key disease-specific parameter in our model, which can also be expressed as the rate of quanta emission by an infected person or in simpler words the number of the viruses exhaled by an infected person.

3. A Case for Study

Whether you are close enough to the carrier or across the room, the risk will be similar.Under these conditions, what can we do?

To answer the question above they will take a concrete case that we should be talking about. Imagine a class of college students, of which one of the students is carrying without knowing it. That it is not very symptomatic, or that he has not yet been diagnosed, they can imagine that he will stay a few days in class, being contagious. Question: how many comrades will he infect in his classroom, and how to make it so that it is if possible 0? To answer this question, they are going to do some small calculations from simple physical principles.There are going to be some formulas.

3.1 Exhaled Air

The first thing to know is that because of its simple breathing, our infected student will reject about $0.5m^3$ of air per hour. If he talks a lot or sings, the values will be higher, up to $1m^3$ of exhaust air per hour. They will call **Q** the flow of exhaled air. Now the question is how many viruses are there in that exhaled air? So it's pretty hard to know, and they'll presume a value. They will say that there are 1000 copies of the virus per m^3 of exhaled air.Obviously they have to keep in mind that these copies of the virus are not in the open air, they are in micro-droplets, aerosols, but their hypothesis is that in all the aerosols contained in $1m^3$ of exhaled air, that's 1000 copies of the virus.They will note that$\boldsymbol{C_v}$, the virus concentration in the exhaled air.

3.2 The Mask

Then there is a super important thing: the mask. If it is well put on, it will absorb a good part of the droplets, therefore copies of the virus. A surgical mask has an efficiency of about 90%, that is to say that only 10% of the viruses passing; they will denote this fraction **f**, so 0.1. By the way they will notice two things: first, it is important that it is a mask and a visor. A visor can capture big droplets that are projected when you sneeze, but aerosols in suspension are not stopped by the visor. The other thing is that the mask must be worn well, positioned on the nose. They are talking about droplets contained in the breath, breathing is through the nose a lot, so if the mask is under the nose, all that contaminated air will come out without a problem. So for their students, they can assume that the mask is not always worn perfectly, and that in total there are 15% of viruses passing, so **f** = 0.15. If they make the product of these 3 quantities: 0.5m^3/ h, 1000 viruses per m^3 and 15% of passing viruses, they obtain the quantity of virus released into the ambient air by their patient each hour. Here **Q** times **f** times$\boldsymbol{C_v}$, that's about 75 expired viruses per hour.

3.3 The Micro-Droplets

Then they must take into account that the micro-droplets which contain the viruses, once rejected, they will quickly diffuse throughout the room. So the 75 viruses will end up diluted throughout the volume of the room. If the classroom is 50m^2 and has a ceiling height of 3m, these 75 viruses will be diluted in a volume **V** of 150m^3 of air. So the average concentration once diluted, it's going to be 0.5 virus per$\boldsymbol{m^3}$. It doesn't seem like a lot. Except remember, that's what's added to the air every hour. So in theory the virus concentration in the ambient air should increase continuously: 0.5, 1,and 2 etc. Hour after hour. Finally that would be if the classroom were a perfectly airtight jar, fortunately a room is never airtight, there are air vents, maybe even mechanical ventilation which extracts the air and rejects it to the room. Outside and probably those we ventilate from time to time by opening the windows.

3.4 Air Renewal Rate

There are several ways to quantify all these phenomena; a simple way is to look at the rate of renewal of the air in the room: how many times the air in the room is renewed each hour. Suppose it takes two hours to renew everything, that means the renewal rate is 0.5 times per hour. They will note that **lambda**. Basically this means that each hour, the viruses released do not dilute in 150m^3 of fresh air, but in 0.5 times this value. By doing the calculation well, they can show that between what is rejected by the patient and what is renewed; the concentration of virus in the ambient air will reach an equilibrium which is equal to **Q** × **f** ×$\boldsymbol{C_v}$, divided by **lambda** × V. With the figures they have chosen, that puts us at 1 virus perm^3. So how much is 1 virus perm^3?

3.5 A Simple Scenario

Well now they're going to try to estimate how many infections it's going to cause. Let's put ourselves in the shoes of another student located elsewhere in the classroom. He will also breathe the ambient air, and he will inhale and exhale at the same rate as what they assumed earlier: a **Q** rate of 0.5 m^3/ h. If the contagious student stays, say, 3 days in class before being tested and isolated, that means that at 7 hours of class per day, their healthy student will be exposed to the ambient virus for about twenty hours in total. They will write "**tau**" this exposure time. About twenty hours at 0.5m^3 per hour, that's 10m^3 of inhaled air.If they multiply that by the concentration of virus in the ambient air, there is one virus per m^3 , so it's makes 10 copies of the virus which will be inhaled by this student over the 3 days of exposure. And except that they forgot something, the mask! Yes it also works in this sense, the mask will filter 90% of the incoming viruses, so only one out of the 10 will pass, that makes only one inhaled virus. The complete formula of all their calculations, the number of viruses inhaled in total is: **tau**$\boldsymbol{Q^2 f^2}$ $\boldsymbol{C_v}$ divided by **lambda V**.

So an inhaled virus, but is it serious? Is that enough to nab the COVID? Well probably not! With each copy of virus, there is a certain probability "**p**" that it really infects you. How much is this probability worth? They will say 10%: only one risk in 10 for each copy of the virus. They will therefore multiply their formula by **P**, and that will give themtheir probability of being infected. Their poor student, who has inhaled on average one virus during these 20 hours of exposure, has only a 10% chance of getting sick. Frankly it's okay! It’s okay, except that there are 30 in the class, to be just as exposed to this mode of contamination, so if each student has a 10% chance, they are pretty sure to recover one, or even several cases. Basically for simplicity, they can multiply their formula by **N, N:** number of people present. And this big formula, it roughly gives the probability that there is a transmission of the virus. And they remind you, even if there is only one transmission, in fact it is a lot. If each patient transmits it to more than one person, the number of reproduction, the famous "R0" will be greater than 1, and therefore the epidemic will continue exponentially.

3.6 The Effectof other Factors

What they want is to bring that below 1. They want on average less than one transmission per patient. And since they are just reasoning about the time spent in class, they would like to bring this figure really below 1. So for that let's look at this formula together, it tells us some interesting things about aerosol transmission. First, it's proportional to the cumulative exposure time **tau**, It's pretty logical, if you stay twice as long, the risk doubles. Then the filtration coefficient **f** appears squared, because there is a double effect of the masks. They limit the viruses expired by the carrier, and those inhaled by the recipient. A well-worn mask only has 10% of the virus passing through each time, so that in total reduces transmission by a factor of 100. The effect of is enormous.

In the denominator, they see that they have the volume of the room and the air renewal rate the larger the room and the more air is renewed, the less risks they take, it is logical, but it is good to see it. And then here they have two terms left. The concentration of virus in the exhaled air: $\boldsymbol{C_v}$, and the probability of contamination **P** when inhaling a copy of the virus. They have given you values, which are credible, but about which they actually have great uncertainty. It's pretty hard to know that. So does that discredit the whole analysis? Well no, because what really matters, you see, is the product of these two quantities. 1000 viruses per m^3expired, and 10% probability of contamination, is the same as 10,000 viruses per $m^3$and 1% probability. Imagine that it is 10%, this probability, that means that it takes on average an exposure to 10 copies of virus to be contaminated, therefore 1000 viruses per m^3, it is 100 times this quantity, this "dose", this dose they will call it: a quantum of infection. They say "quantum" but nothing to do with quantum mechanics; it's just the typical amount it takes to get infected. And their 1000 viruses per$\boldsymbol{m^3}$, thus representing 100 quantum of infection per$\boldsymbol{m^3}$. And if it had been 10,000 viruses with a probability of 1%, it would have been the same thing, it would also have made 100 quantum of infection per$\boldsymbol{m^3}$. So in their analysis, they will forget the virus concentration and the probability of infection, they are going to replace by the product of the two, the concentration in quantum of infection, they will note it $\boldsymbol{C_q}$, and why is it better?

3.7 Concentration in quantum of infectionC_q

Because it is something that can be more easily estimated. Martin Bazant and John Bush, the physicists who wrote the article, took data from various contamination events, and were able to estimate the mean values for the quantum concentration of infection emitted by a patient. The typical value obtained for normal activity is around 70 quantum per $m^3$of exhaled air. But for the case of the choir, where people were singing, the value would be closer to 900 quantum perm^3. They can easily imagine that when we sing loudly, we expel more droplets, which may come from different regions of the lungs, and that this greatly increases the concentration of infection quantum in the exhaled air. Obviously this estimate will depend on a whole bunch of factors: the patient's symptom state, the target population and of course, the strain of the virus. While some variants, such as the famous English variant, are more contagious, this will result in higher quantum concentrations. Note that they do not necessarily know if with the English variant the aerosol droplets are more concentrated in virus, or if it is the probability of infection associated with a copy of the virus which is higher, but the effect on the infection quantum concentration is the same, it will increase it maybe by 50%. In any case, you can see that the value of 100 quantum per $m^3$was quite reasonable, they don't have a super precise value, but it is essentially the only parameter of the analysis on which they have an uncertainty.

By taking a good safety margin it can allow us to advance in the reasoning. So let's come back to our formula, they said that they wanted to limit the reproduction of the virus, and therefore to lower this value sufficiently below

1. To be broad with their uncertainties, they will say that they set their selves a tolerance threshold at 10%, they will note it **epsilon** and so here is what they want: all that less than **epsilon**. What can they do to meet this threshold?

3.8 The Measures needed

They can ask their selves the question for each situation where people are brought together in the same room. There they will look with the characteristics that they gave for the classroom. First, they can limit the number of people present **N** and the exposure time **tau**. From this point of view, the practice of the half-gauge is obviously going in the right direction. Another essential thing, you have to wear the mask well, they remind you that the factor **f** which has a squared effect. It is also necessary to avoid activities such as singing or sport which increases the rate of respiration **Q**. For sport, it can add a factor of 10 to the rate, squared that makes 100. So indoor sport and worse, without a mask, it's a no! The volume of room **V**, a priori they cannot change it too much. But they still have **lambda**, the air renewal rate. They have been quite pessimistic, they have taken 0.5: a renewal every 2 hours. This is a typical value for residential premises, but in collective premises, with a good ventilation system or periodic ventilation, much better can be done. They can multiply this value by 5 or even 10, and therefore reduce the risk accordingly. If you have your hand on a mechanical ventilation system, a typical benchmark is to impose a ventilation rate of about $30m^3$ per hour and per person. With 30 students, that's 900 m^3per hour, so 150 $m^3$of classroom will be renewed 6 times per hour. 6 times for **lambda** instead of 0.5 the risk was reduced by a factor of 12. That's good if there is mechanical ventilation whose flow is controlled, but it is likely that in many situations this is not the case. [1]

Exhaled air flow: **Q**

Fraction that crosses the mask: **f**

Room volume: **V**

Air Renewal rate:**λ**

Exposure time: **t**

Number of persons: **N**

Quantum concentration of infection in exhaled air:$\boldsymbol{C_q}$

Total probability of transmission:$\frac{\mathbf{N}*\mathbf{t}*\,Q^2*f^2*\boldsymbol{C_q}}{\boldsymbol{\lambda}\mathbf{V}}$

4. Explanation

In order to create a tool of estimation we are going to incorporate and shape reinforcement learning around the formula proposed by [1].

So what was our contribution? What we did was we took that formula and we added the 3 past weeks estimation plus the current week to estimate the risk for the next month and the month after that. And by doing that we are teaching our agent to change the value of $\boldsymbol{C_q}$depending on the progress of the estimation during the monthso how it works:

In reinforcement learning we have a decision maker called an agent that interacts with the environment that it's placed in these interactions occur sequentially over time at each time step the agent will get some representation of the environment state and given this representation that agent select an action to take the environment is then transitioned into some new state and the agent is getting a reward as a consequence of its previous action so to summarize the components of the reinforcement learning model include the environment the agent all the possible states of the environment all the action that the agent can take in the environment and all the rewards that the agent can receive from taking actions in the environment so the environment for our agent are the inputs:

Number of persons: **N**

Exposure time: **t**

Exhaled air flow: **Q**

Fraction that crosses the mask: **f**

Quantum concentration of infection in exhaled air: $\boldsymbol{C_q}$

Air Renewal rate: **λ**

Room volume: **V**

And then the past three weeks estimation transmission (S1,S2,S3) plus current week estimation **rp.**

The agent will receive these inputs and then calculate the estimation of transmission pc and the total probability of transmission tp, then the difference between the past three weeks and the current week. By doing all this calculations the agent now will have the state of the environment as it shown in the caption below (I is the number of weeks in the month)

```
s1=input("First Week Transmission Estimation: ") ; s1=float(s1)
s2=input("Second Week Transmission Estimation: ") ; s2=float(s2)
s3=input("Third Week Transmission Estimation: ") ; s3=float(s3)
rp=input("Current Week Transmission Estimation: ") ; rp=float(rp)
n=input("Number of persons: ") ; n=float(n)
t=input("Exposure time: ") ; t=float(t)
q=input("Exhaled air flow: ") ; q=float(q)
f=input("Fraction that crosses the mask: ") ; f=float(f)
cq=input("Quantum concentration of infection in exhaled air: ") ; cq=float(cq)
λ=input("Air Renewal rate: ") ; λ=float(λ)
```

```
v=input("Room volume: ") ; v=float(v)
pc=(n*t*(q*q)*(f*f)*cq)/(λ*v)
tp=((((s1+2*(s2)+3*(s3))/6)+(4*rp))/5)-pc
print("The Old Total Transmission Estimation befor the RL Algorithm:",pc)
print("The Old Total probability of transmission befor the RL Algorithm:",tp)
i=1
```

The states are **pc**(Total Transmission Estimation) and **tp**(Total probability of transmission) depending on these states the agent will select an action and the action for our agent is to decide $\boldsymbol{C_q}$ is positive or negative if positive there will be a reward if negative a penalty as it shown in the caption below

```
whilei<= 4:
 iftp<=-0.03:
   cq=cq-0.3
   pc=(n*t*(q*q)*(f*f)*cq)/(λ*v)
   tp=((((s1+2*(s2)+3*(s3))/6)+(4*rp))/5)-pc
   print("Total Estimation of transmission For The",i,"Week After Adjusting is:",pc)
   print("Total probability of transmission For The",i,"Week After Adjusting is:",tp)
   s1=s2
   s2=s3
   s3=pc
 eliftp>-0.03andtp<=-0.02:
       cq=cq-0.2
       pc=(n*t*(q*q)*(f*f)*cq)/(λ*v)
       tp=((((s1+2*(s2)+3*(s3))/6)+(4*rp))/5)-pc
       print("Total Estimation of transmission For The",i,"Week After Adjusting is:",pc)
       print("Total probability of transmission For The",i,"Week After Adjusting is:",tp)
       s1=s2
       s2=s3
       s3=pc
 eliftp>-0.02andtp<=0:
   cq=cq-0.1
   pc=(n*t*(q*q)*(f*f)*cq)/(λ*v)
   tp=((((s1+2*(s2)+3*(s3))/6)+(4*rp))/5)-pc
   print("Total Estimation of transmission For The",i,"Week After Adjusting is:",pc)
   print("Total probability of transmission For The",i,"Week After Adjusting is:",tp)
```

```
  s1=s2
  s2=s3
  s3=pc
eliftp>0andtp<=0.02:
  cq=cq+0.1
  pc=(n*t*(q*q)*(f*f)*cq)/(λ*v)
  tp=((((s1+2*(s2)+3*(s3))/6)+(4*rp))/5)-pc
  print("Total Estimation of transmission For The",i,"Week After Adjusting is:",pc)
  print("Total probability of transmission For The",i,"Week After Adjusting is:",tp)
  s1=s2
  s2=s3
  s3=pc
eliftp>0.02andtp<=0.03:
  cq=cq+0.2
  pc=(n*t*(q*q)*(f*f)*cq)/(λ*v)
  tp=((((s1+2*(s2)+3*(s3))/6)+(4*rp))/5)-pc
  print("Total Estimation of transmission For The",i,"Week After Adjusting is:",pc)
  print("Total probability of transmission For The",i,"Week After Adjusting is:",tp)
  s1=s2
  s2=s3
  s3=pc
eliftp>0.03:
  cq=cq+0.3
  pc=(n*t*(q*q)*(f*f)*cq)/(λ*v)
  tp=((((s1+2*(s2)+3*(s3))/6)+(4*rp))/5)-pc
  print("Total Estimation of transmission For The",i,"Week After Adjusting is:",pc)
  print("Total probability of transmission For The",i,"Week After Adjusting is:",tp)
  s1=s2
  s2=s3
  s3=pc
else:
  print("null")
i = i + 1
```

This process of selecting an action from a given state transitioning to a new state and receiving a reward happens sequentially over and over again (until i=4 witch means a whole month) which creates something called a trajectory that shows the sequence of state actions and rewards throughout the process, so there are a set of states S, a set of actions A, a set of rewards R.

At each time step t=0,1,2,......, the agent receives some representation of the environment's state $\boldsymbol{S_t}$. Based on this state, the agent select an action $\boldsymbol{A_t}$. This gives us the state-action pair ($\boldsymbol{S_t}$,$\boldsymbol{A_t}$). Time then is incremented to the next time step t+1 and the environment is transitioned to a new state St+1. $\boldsymbol{A_t}$this time the agent receives a numerical reward Rt+1 for the action $\boldsymbol{A_t}$taken from the state $\boldsymbol{S_t}$.

So in simpler way

Step 1

At time t the environment is in state $\boldsymbol{S_t}$.

Step 2

The agent observes the current state and select action $\boldsymbol{A_t}$.

Step 3

The environment transitions to state $\boldsymbol{S_t}$+1 and grants the agent reward $\boldsymbol{R_t}$+1.

This process then starts over for the next time step T+1 until i=4

So the policy for this agent is plain and simple calculate the estimation and then the difference depending on the difference there will be either a reward or a penalty.Also at the same time by using the estimation of the previous month we’ll be able to estimatethe probability oftransmissionfor the next 4 weeks.

5. Integration of RL

Begin

Read(S1,S2,S3,n,t,q,f,C_q,λ,v,rp);// read the inputs entered by the user

Pc= (n*t*(q*q)*(f*f)*C_q)/(λ*v);// calculate the estimation of transmission

Tp= ((((S1+2*(S2)+3*(S3))/6)+(4*rp))/5)-pc;// calculate the total probability of transmission or in other words the difference between past 3 weeks and currensst

Write (“The transmission estimation:”,pc);

Write (“The total probability of transmission:”,tp);

i=1;

while i<=4// depending on the value of tp our agent is learning how to chang the value C_q

Begin

If tp<=-0.03// in this case our agent will get 0.3 penalty forC_q

Begin

then cq=cq-0.3; recalculate pc and tp; s1=s2; s2=s3; s3=pc;

Write (“The transmission estimation:”,pc);

Write (“The total probability of transmission:”,tp);

Else If tp>-0.03 and tp<=-0.02// in this case our agent will get 0.2 penalty forC_q

then cq=cq-0.2; recalculate pc and tp; s1=s2; s2=s3; s3=pc;

Write (“The transmission estimation:”,pc);

Write (“The total probability of transmission:”,tp);

Else If tp>-0.02 and tp<=0// in this case our agent will get 0.1 penalty forC_q

then cq=cq-0.1; recalculate pc and tp s1=s2; s2=s3; s3=pc;

Write (“The transmission estimation:”,pc);

Write (“The total probability of transmission:”,tp);

Else If tp>0 and tp<=0.02// in this case our agent will get 0.1 reward forC_q

then cq=cq+0.1 recalculate pc and tp; s1=s2; s2=s3; s3=pc;

Write (“The transmission estimation:”,pc);

Write (“The total probability of transmission:”,tp);

Else If tp>0.02 and tp<=0.03// in this case our agent will get 0.2 reward forC_q

then cq=cq+0.2; recalculate pc and tp; s1=s2; s2=s3; s3=pc;

Write (“The transmission estimation:”,pc);

Write (“The total probability of transmission:”,tp);

Else If tp>0.03// in this case our agent will get 0.3 reward forC_q

then cq=cq+0.3 recalculate pc and tp; s1=s2; s2=s3; s3=pc;

Write (“The transmission estimation:”,pc);

Write (“The total probability of transmission:”,tp);

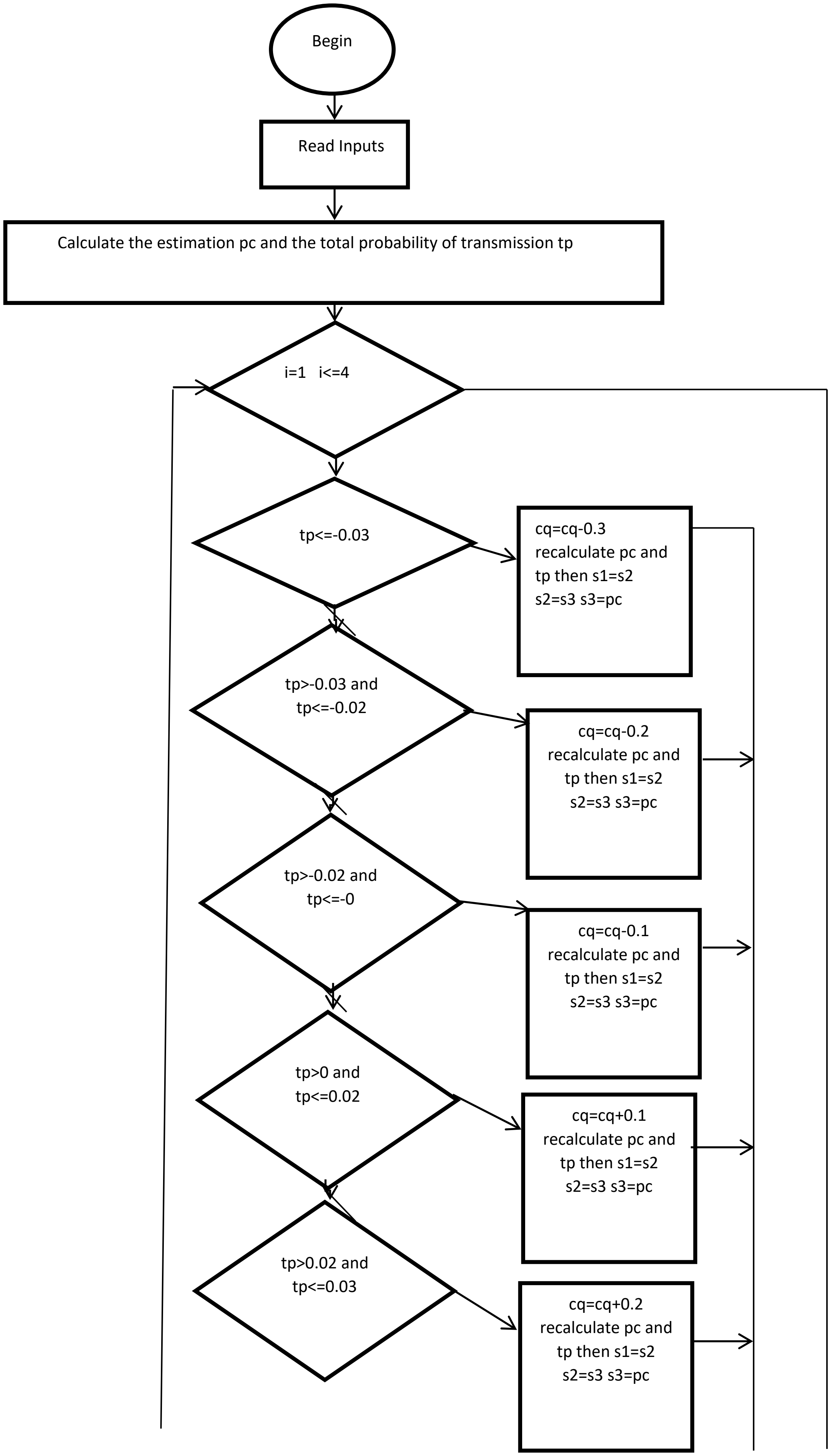
Else Write (“Null”);

Endif;

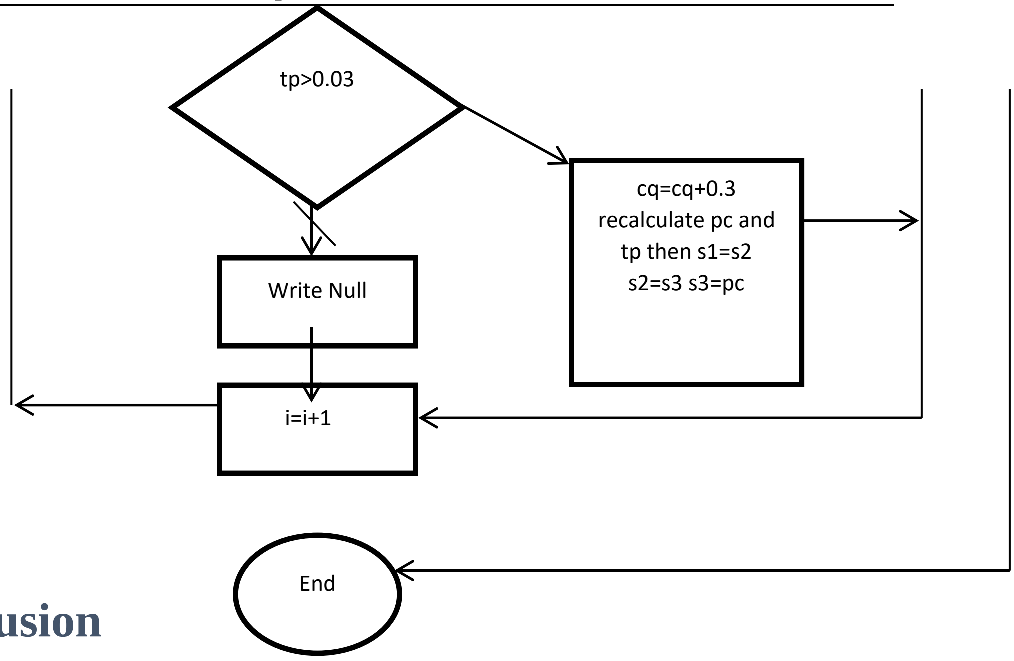
i=i+1

End;

6. AlgorithmOrganizational chart



Begin
Read Inputs
Calculate the estimation pc and the total probability of transmission tp
i=1 i<=4
tp<=-0.03
cq=cq-0.3 recalculate pc and tp then s1=s2 s2=s3 s3=pc
tp>-0.03 and tp<=-0.02
cq=cq-0.2 recalculate pc and tp then s1=s2 s2=s3 s3=pc
tp>-0.02 and tp<=-0
cq=cq-0.1 recalculate pc and tp then s1=s2 s2=s3 s3=pc
tp>0 and tp<=0.02
cq=cq+0.1 recalculate pc and tp then s1=s2 s2=s3 s3=pc
tp>0.02 and tp<=0.03
cq=cq+0.2 recalculate pc and tp then s1=s2 s2=s3 s3=pc



7. Conclusion

In this chapter we discussed the integration of RL and the implementation and now remain the tests and results which we will do in the next chapter.

Chapter 4 Tests and Results

1. Development Enviroment

1.1 Choice of Language: PYTHON

Our choice of programming language fell on PYTHON 3.7.2, and that's because Python is the most popular language in the world of artificial intelligence. Python is object oriented and is meant to be relatively easy to access. It is widely used within the scientific community and particularly in the field of artificial intelligence.

1.2 ANACONDA

Anaconda is a free and open source distribution of the Python and R programming languages applied to the development of applications for data science and machine learning.

Anaconda will simplify our task since it installs all the packages necessary for machine learning, for example: NumPy, panda, scikit-learn.

1.3 Visual Studio Code (VS code)

Visual Studio Code is an extensible code editor developed by Microsoft.

2. Tests and Scenarios

We start by testing four COVID-19 spreading events by using existing estimates of relevant physical parameters that caused the super-spreading

scenario	N	t	Q	f	C_q	**λ**	**V**
Skagit Church Choir	61	2.5	1.0	0.1	870	0.65	810
Ningbo Tour Bus	68	1.7	0.5	0.1	90	1.25	45
Diamond Princess	3711	288	0.5	0.1	30	8	291900
Wuhan City Outbreak	3.03	132	0.5	0.1	29	0.34	216

Table 2 : Scenarios Table

NisNumber of persons (3.03 in the table above is the average family number that lives in each appartment)

t isExposure time

Exhaled air flow: **Q**

f isFraction that crosses the mask

C_qIsQuantum concentration of infection in exhaled air

λisAir Renewal rate

VisRoom volume

Scenario 1: skagit Church Choir

The Skagit Valley Choir event we use existing estimates of relevant physical parameters.

Results

We notice that the estimation of transmission of the previous month is: 2.51 which mean that each person will spread the virus to another 2.51 people so if we have a group of 10 people they'll contaminate another 25 people, and the probability of transmission is 0.19 in other words there is a chance of 19% of contamination.

Also the estimation of transmission of the 1 week is: 2.52

The probability of transmission of the 1 week is: 0.194

The estimation of transmission of the 2 week is: 2.521

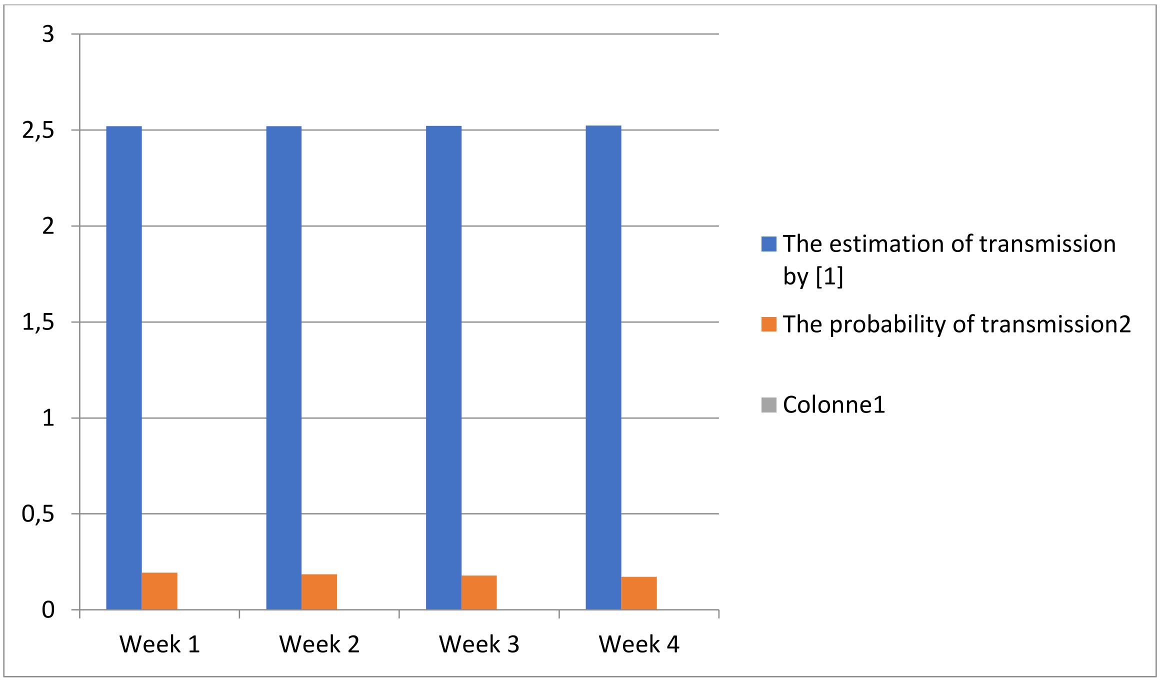
The probability of transmission of the 2 week is: 0.186

The estimation of transmission of the 3 week is: 2.522

The probability of transmission of the 3 week is: 0.178

The estimation of transmission of the 4 week is: 2.523

The probability of transmission of the 4 week is: 0.172



Results Discussion

We also notice that for the next 4 weeks the estimation is getting higher while the opposite for the probability because our agent is adding to C_q which means higher transmission rate all this that the difference between the estimation and the probability of previous month is not that big.

We also notice that the risque in this scenario is so high due to the outstanding C_q value because of all the singing and the shouting.

Scenario 2: ningbo Tour Bus

A tour bus transported 68 people (including the driver) on a 100 minute round-trip journey to a Buddhist ceremony in Ningbo, China. One index case infected 23 fellow passengers, three of which are assumed to have been infected at the ceremony.

Results

We notice that the estimation of transmission of the previous month is: 0.462 which mean that each person will spread the virus to another 0.46 people so if we have a group of 10 people they'll contaminate another 4.6 people and the probability of transmission is 0.051 in other words there is a chance of 5.1% of contamination.

Also the estimation of transmission of the 1 week is: 0.463

The probability of transmission of the 1 week is: 0.05

The estimation of transmission of the 2 week is: 0.465

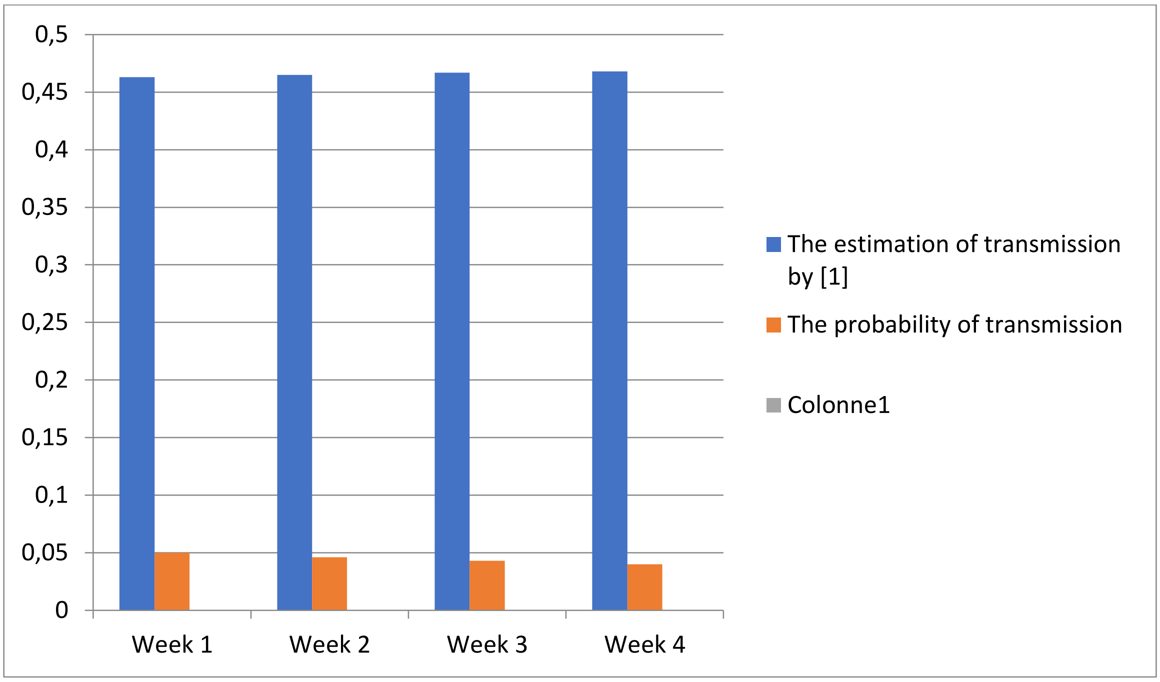
The probability of transmission of the 2 week is: 0.046

The estimation of transmission of the 3 week is: 0.467

The probability of transmission of the 3 week is: 0.043

The estimation of transmission of the 4 week is: 0.468

The probability of transmission of the 4 week is: 0.04



Results Discussion

We also notice that for the next 4 weeks the estimation is getting higher while the opposite for the probability because our agent is adding to C_q which means higher transmission rate all this that the difference between the estimation and the probability of previous month is not that big.

We also notice that the risque in this scenario is high due to the small space in the bus.

Scenario 3: diamond Princess

The Diamond Princess cruise ship during where passengers and crew mainly occupy 14 floors of living space of beam width 38 m, an average length equal to 90% of the ship's length 290 m.

Results

We notice that the estimation of transmission of the previous month is: 0.343 which mean that each person will spread the virus to another 0.343 people so if we have a group of 10 people they'll contaminate another 3.43 people and the probability of transmission is 0.0112 in other words there is a chance of 5.1% of contamination.

Also the estimation of transmission of the 1 week is: 0.344

The probability of transmission of the 1 week is: 0.0111

The estimation of transmission of the 2 week is: 0.345

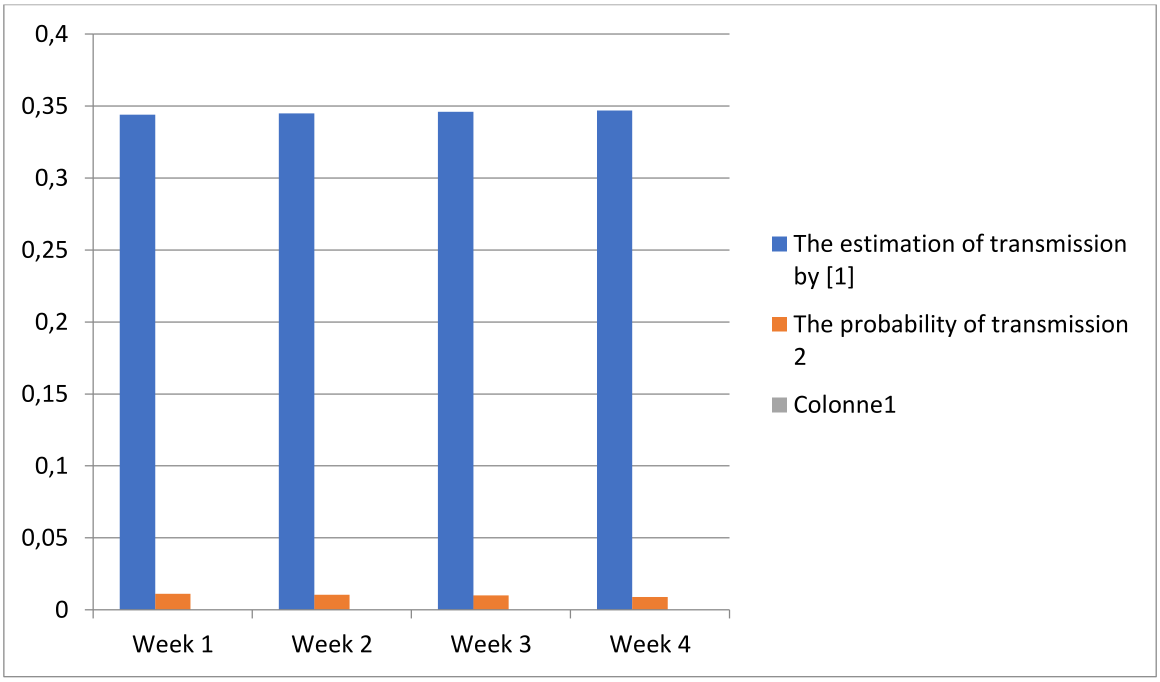
The probability of transmission of the 2 week is: 0.0105

The estimation of transmission of the 3 week is: 0.346

The probability of transmission of the 3 week is: 0.010

The estimation of transmission of the 4 week is: 0.347

The probability of transmission of the 4 week is: 0.009



Results Discussion

We also notice that for the next 4 weeks the estimation is getting higher while the opposite for the probability because our agent is adding to C_q which means

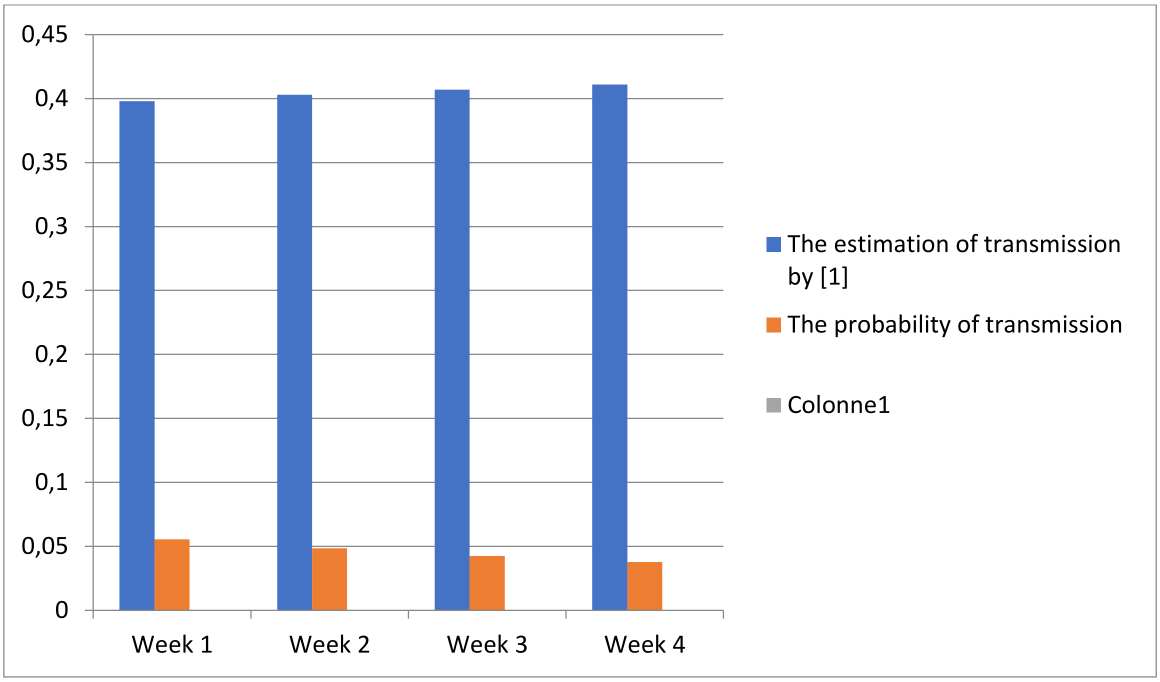
higher transmission rate all this that the difference between the estimation and the probability of previous month is not that big.
We also notice that the risque in this scenario is not that high even with the enormous number of the people aboard and the exposer time (288 hours) it’s because the space was too big and the higher air renewal rate.

Scenario 4: wuhan City Outbreak

Initial outbreak in Wuhan City, Hubei Province, China. We assume that the population-level spreading is dominated by indoor aerosol transmission with slow incubation in single-family apartments with a mean family size of 3.03, in mean apartment area of $216m^3$.

Results

We notice that the estimation of transmission of the previous month is: 0.394 which mean that each person will spread the virus to another 0.394 people so if we have a group of 10 people they’ll contaminate another 3.94 people and the probability of transmission is 0.0594 in other words there is a chance of 5.1% of contamination.
Also the estimation of transmission of the 1 week is: 0.398
The probability of transmission of the 1 week is: 0.0554
The estimation of transmission of the 2 week is: 0.403
The probability of transmission of the 2 week is: 0.0485
The estimation of transmission of the 3 week is: 0.407
The probability of transmission of the 3 week is: 0.0424
The estimation of transmission of the 4 week is: 0.411
The probability of transmission of the 4 week is: 0.0376



Results Discussion

We also notice that for the next 4 weeks the estimation is getting higher while the opposite for the probability because our agent is adding to C_q which means higher transmission rate all this that the difference between the estimation and the probability of previous month is not that big.

We also notice that the risque in this scenario is high because of the low air renewal rate and the relatively small apartment with the exposure time of 132 hours and another factor not mentioned the fact that the ventilation system of this apartment is connected to all other apartment which means that one person will contaminate entire building.

Scenario 5: University Of Saad Dahleb Blida

For the next test we'll go to a familiar place to our own department of computer science more specifically to the SIR students of which am one of them.

We'll have two cases

CASE 1

For the lessons they'll have us all studying in a relatively larger room then usual so V=120m^3 the students present plus the teacher N=62, Q=0.5, t=4 hours (during pandemic), f=0.1, λwe'll set it for 0.5, C_q= most of the students will be doing nose to nose breathing only the teacher will be explaining the lesson so we'll set it to 10.9

Results

We notice that the estimation of transmission of the previous month is: 0.112 which mean that each person will spread the virus to another 0.112 people so if we have a group of 10 people they'll contaminate another 1.12 people and the probability of transmission is 0.00063 in other words there is a chance of 0.063% of contamination.

Also the estimation of transmission of the 1 week is: 0.111

The probability of transmission of the 1 week is: 0.0003

The estimation of transmission of the 2 week is: 0.112

The probability of transmission of the 2 week is:-0.0011

The estimation of transmission of the 3 week is: 0.111

The probability of transmission of the 3 week is: -0.0008

The estimation of transmission of the 4 week is: 0.110

The probability of transmission of the 4 week is: -0.0001

Results Discussion

As we can see the risk isn't that big at all because of the big volume of the class so the bigger the room the more air is renewed the lower the risk.

CASE 2

For the TD it's pretty much the same only

V=60m^3, C_q=14.5, with N=23.

Results

We notice that the estimation of transmission of the previous month is: 0.111 which mean that each person will spread the virus to another 0.111 people so if we have a group of 10 people they'll contaminate another 1.11 people and the probability of transmission is 0.026 in other words there is a chance of 2.6% of contamination.

Also the estimation of transmission of the 1 week is: 0.112

The probability of transmission of the 1 week is: 0.024

The estimation of transmission of the 2 week is: 0.114

The probability of transmission of the 2 week is: 0.022

The estimation of transmission of the 3 week is: 0.115

The probability of transmission of the 3 week is: 0.020

The estimation of transmission of the 4 week is: 0.117

The probability of transmission of the 4 week is: 0.018

Results Discussion

As we can see the risk is bigger because of the smaller class with a higherQuantum concentration of infection in exhaled air.

3. Conclusion

From the scenarios above the measures needed are simple, we can limit the number of people present N and the exposure time tau. From this point of view, the practice of the half-gauge is obviously going in the right direction. Another essential thing, you have to wear the mask well, I remind you that the factor f which has a squared effect. It is also necessary to avoid activities such as singing or sport which increases the rate of respiration Q. For sport, it can add a factor of 10 to the rate, squared that makes 100. So indoor sport and worse, without a mask, it's a no! The volume of room V, a priori we cannot change it too much. But we still have lambda, the air renewal rate. They have been quite pessimistic, they have taken 0.5: a renewal every 2 hours. This is a typical value for residential premises, but in collective premises, with a good ventilation system or periodic ventilation, much better can be done. We can multiply this value by 5 or even 10, and therefore reduce the risk accordingly.

General conclusion

Nowadays, it is observed that this mode of transmission of the virus indoor airborne disease transmission plays a very important role in the virus breakthrough.we have the number of people present N and the exposure time tauQuantum concentration of infection in exhaled air$\boldsymbol{C_q}$Exhaled air flow: $\mathbf{Q}$ which they all have a significant influence in this mode of transmission.

We have managed, through this project, to provide a quantitative analysis of the phenomenon, in order to have guides on the attitude to adopt. The goal is to limit in a reasonable way this mode of transmission of the virus based on Reinforcement learning which consists of an state $\boldsymbol{S_t}$ layer action $\boldsymbol{A_t}$ layer, an reward $\boldsymbol{R_t}$layer. Our model has allowed for an estimate of the infectiousness of COVID-19, it also makes clear the inadequacy of the Six-Foot Rule in mitigating indoor airborne disease transmission, and offers a rational, physically informed alternative for managing life in the time of COVID-19. For each situation where people are brought together in the same room.

This work was an opportunity for me to complete my computer skills in a transversal way and to broaden and deepen my knowledge of artificial intelligence. However, prospects for improving our model remain conceivable to be enriched by to creating an app not only to estimate of the infectiousness of COVID-19 also to have guides on the attitudes to adopt in each situation.

References

1. Martin Z. Bazant and John W. M. Bush,A guideline to limit indoor airborne transmission of COVID-19, 2019.
2. Dr. Hentabli hamza, Deep Learning, Supervisor Prof. Naomie Salim, 2019.
3. Robert Moni,Reinforcement Learning algorithms an intuitive overview 2019.
4. Kung-Hsiang, Huang (Steeve), Introduction to Various Reinforcement Learning Algorithms, 2018.
5. Aditya Gudimella, RossStory, MatinehShaker, Ruofan Kong, Matthew Brown, VictorShnayder, Marcos Campos, Concept networks for grasp & stack, 2017.
6. Peter Dayana and Yael Niv — Reinforcement learning: The Good, The Bad and The Ugly, 2008.
7. J. A. Lednicky *et al.*, Viable SARS-CoV-2 in the air of a hospital room with COVID-19, 2020.
8. M. Jayaweera, H. Perera, B. Gunawardana, J. Manatunge, Transmission of COVID-19 virus by droplets and aerosols, 2020.
9. R. Mittal, R. Ni, J.-H. Seo, The flow physics of COVID-19, 2020.
10. L. Morawksa, Droplet fate in indoor environments, or can we prevent the spread of infection? ,2006.
11. L. Morawska *et al.*, Size distribution and sites of origin of droplets expelled from the human respiratory tract during expiratory activities, 2009.
12. J. A. Lednicky *et al.*, Viable SARS-CoV-2 in the air of a hospital room with COVID-19, 2020.
13. M. Jayaweera, H. Perera, B. Gunawardana, J. Manatunge, Transmission of COVID-19 virus by droplets and aerosols, 2020.
14. R. Mittal, R. Ni, J.-H. Seo, The flow physics of COVID-19, 2020.
15. L. Morawksa, Droplet fate in indoor environments, 2006.
16. L. Morawska *et al.*, Size distribution and sites of origin of droplets expelled from the human respiratory tract during expiratory activities, 2009.
17. L. Morawska, D. K. Milton, It is time to address airborne transmission of COVID-19, 2020.
18. L. Morawska, J. Cao, Airborne transmission of SARS-CoV-2: The world should face the reality, 2020.
19. M. Jayaweera, H. Perera, B. Gunawardana, J. Manatunge, Transmission of COVID-19 virus by droplets and aerosols, 2020.
20. G. Buonanno, L. Stabile, L. Morawska, Estimation of airborne viral emission: Quanta emission rate of SARS-CoV-2 for infection risk assessment, 2020.
21. Setti *et al.*, Airborne transmission route of COVID-19: Why 2 meters/6 ft of interpersonal distance could not be enough, 2020.
22. R. Zhang, Y. Li, A. L. Zhang, Y. Wang, M. J. Molina, Identifying airborne transmission as the dominant route for the spread of COVID-19, 2020.
23. S. L. Miller *et al.*, Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the Skagit Valley Chorale superspreading event, 2020.

24. Y. Shen *et al.*, Community outbreak investigation of SARS-CoV-2 transmission among bus riders in eastern China, 2020.
25. H. Nishiura *et al.*, Closed environments facilitate secondary transmission of coronavirus disease, 2019 (COVID-19).
26. L. Hamner, High SARS-CoV-2 attack rate following exposure at a choir practice, Skagit County, Washington, March 2020.
27. L. Gammaitoni, M. C. Nucci, Using a mathematical model to evaluate the efficacy of TB control measures. *Emerg. Infect. Dis.* **3**, 335 (1997).
28. C. B. Beggs, C. J. Noakes, P. A. Sleigh, L. A. Fletcher, K. Siddiqi, The transmission of tuberculosis in confined spaces: An analytical review of alternative epidemiological models, 2003.
29. E. C. Riley, G. Murphy, R. L. Riley, Airborne spread of measles in a suburban elementary school, 1978.
30. C.-M. Liao, C.-F. Chang, H.-M. Liang, A probabilistic transmission dynamic model to assess indoor airborne infection risks, 2005.
31. S. N. Rudnick, D. K. Milton, Risk of indoor airborne infection transmission estimated from carbon dioxide concentration, 2003.
32. Impact of COVID-19 on people's livelihoods, their health and our food systems 13 October 2020.
33. Artificial Intelligence and the future of global health, 2014.
34. AI and control of Covid-19 coronavirus, 2020.