SAS® GLOBAL FORUM 2021

Paper 1096-2021

The Potential of Simulation Technologies: Multi-Agent Simulation and Reinforcement Learning

Satoki Fujita, Shogo Miyazawa, Ryo Kiguchi, Yuki Yoshida,
Katsunari Hirano and Yoshitake Kitanishi, Shionogi & Co., Ltd.;

ABSTRACT

The society in which we live is constantly changing, as represented by the spread of Covid-19, which occurred in 2020 and still casts a shadow over the world. It is necessary to anticipate such changes in advance and take the best response promptly, and in such cases, various simulation techniques are useful. Among them, Multi-Agent Simulation (MAS), which is assembled from micro factors and reproduces macro phenomena by their interaction, has a high degree of freedom and can flexibly handle detailed situation settings. This paper introduces the attractiveness of it and show that by introducing Reinforcement Learning (RL) there, it is possible to search for optimal decision-making based on the phenomena reproduced by MAS. The future is predicted through simulations (MAS etc.), and the optimum response policy is derived by RL. As a first step to realize such a flow in actual tasks, we show the potential of these technologies through an example of virtual infectious disease spread. We implement MAS and RL using the Python client on the SAS® Viya® to take advantage of a powerful computing environment.

INTRODUCTION

"What if we do ~?" When thinking about a certain hypothesis or measure, there are many situations where we want to reproduce the situation in which they are applied and observe the phenomena that can be derived as a result. The most reliable manner is to observe it through a demonstration in a real world. However, it is often difficult because it takes a lot of time, money, and labor, or it is ethically impossible to conduct studies In such cases, we have seen through those by using "simulations" instead. In this era when there are abundant computer resources, it is efficient and important to go through the cycle that each phenomenon is formulated and reproduced in the virtual space "Theoretical World" on the computer, and the knowledge gained there is utilized in the "Real World" (Figure 1).

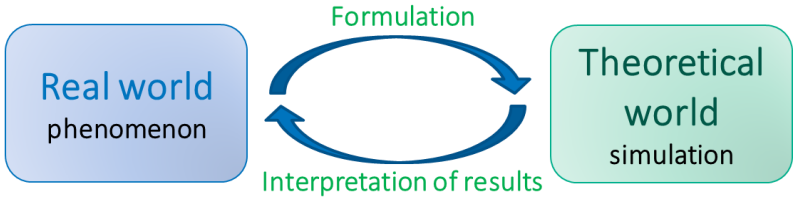


Figure 1. Simulation cycle

However, there are many situations where it is difficult to incorporate the entire phenomenon into the model because various factors are intertwined with each other. For example, the phenomenon of infectious disease spread. In addition to the various states and possible actions of each person (such as the presence or absence of facial coverings, whether or not teleworking etc.), it is difficult to see through their mutual influence. Multi-Agent Simulation (MAS) is a technique that can reproduce such a phenomenon relatively flexibly. MAS is a method of reproducing the phenomenon caused by mutual influence of

micro components by modeling each of them instead of the entire phenomenon. It is used in various fields such as electricity market modeling [1], evacuation route design [2], and ant nest reproduction [3]. Since it is only necessary to formulate each element that composes the target phenomenon, even if the entire phenomenon is complicated, it can be reproduced flexibly.

If the phenomenon can be reproduced by MAS in this way, it is likely that we would like to find the optimal intervention for the phenomenon on the simulation. Because it is a virtual space, you can freely add various changes and observe the results, which is a great merit of simulation. At that time, if there are at most several strategic candidates for interventions, we can try all of them and adopt the one that is in the most ideal state. However, it is not a realistic method if there are innumerable ways, such as when they consist of several combinations. This time, we consider the introduction of Reinforcement Learning (RL) as a countermeasure for such cases. RL is a type of machine learning method that learns the optimal policy by repeating trial and error, and due to its nature, it is compatible with simulations that we can try as many times as we like in virtual space. Based on the above, by combining multi-agent simulation that can flexibly reproduce various phenomena and reinforcement learning that learns from simulation results and automatically finds the optimum action, it may be possible to realize optimal decision-making automation under various situations that can occur in the real world (For example, under a pandemic, derivation of the optimum timing for lockdown in each specific region and optimum strategy for supplying daily necessities to each store, etc.). Here, as a first step for that, we will introduce the application of MAS and RL to virtual cases of infectious disease spread, and show the usefulness and possibility of simulation technology that is expected to be active in various situations in the future.

In order to take the spread of infectious diseases as an example of application of simulation technology, Chapter 1 describes the SEIR model, which is a standard model in the field of infectious diseases. In Chapter 2, based on the concept of the SEIR model, the spread of infectious diseases is reproduced by MAS. Chapter 3 presents an example of searching for an optimal intervention strategy for preventing the spread of infection by RL on the environment simulated in Chapter 2. Finally, a conclusion is given.

1. SEIR MODEL

In this chapter, we will introduce the SEIR model, which is a standard model for predicting the spread of infection, in order to follow this framework later.

In the SEIR model, people are assigned to four states, “Susceptible”, “Exposed”, “Infectious”, and “Recovered”, and the transitions between those states are expressed by ordinary differential equations. Susceptible state refers to a condition that has not yet been infected with a virus and may be infected in the future. Exposed state refers to a state in which he is exposed to the virus but it is latent and has not yet developed as a symptom. Infectious state refers to a condition in which symptoms appear a few days after exposure to the virus. Recovered refers to the state of recovering from an infectious disease and acquiring immunity.

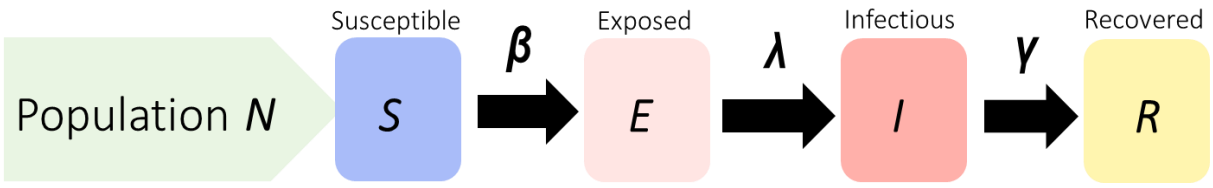


Figure 2. SEIR model

The equations for the transition between each state are as follows.

$$\frac{dS(t)}{dt} = -\beta S(t)I(t)$$
$$\frac{dE(t)}{dt} = \beta S(t)I(t) - \lambda E(t)$$
$$\frac{dI(t)}{dt} = \lambda E(t) - \gamma I(t)$$
$$\frac{dR(t)}{dt} = \gamma I(t)$$

β can be interpreted as the infection rate, $1/\lambda$ as the average latency period, and $1/\gamma$ as the average period from onset to recovery. The initial number of people in each state and the values of the parameter (β, λ, γ) determine the aspect of the infection spread process, and for example, the transition situation shown in the Figure 3 below can be simulated.

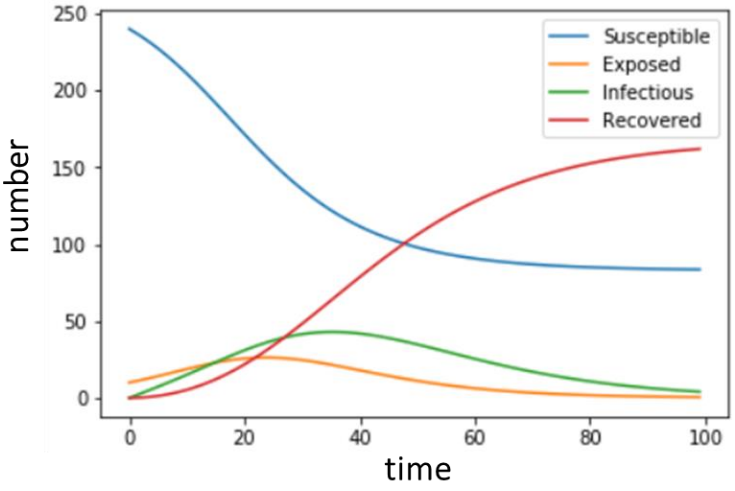


Figure 3. Example of SEIR model plot

This model is very simple and easy to handle because it considers the nature and movement of people to be the same in each of the four states and considers them collectively. However, in reality, people have different backgrounds, thoughts and actions. In the situation where there are those individual differences, it remains doubtful whether the phenomenon of infection spread caused by them can be expressed by the original simple SEIR model. In such cases, Multi-Agent Simulation is useful because it is not necessary to incorporate the entire phenomenon into the model and the individual differences of the components can be flexibly incorporated. It will be described in detail in the next chapter.

2. MULTI AGENT SIMULATION

In this chapter, Multi-Agent Simulation (MAS) is explained by taking the spread of infectious diseases as an example.

2.1. SIMPLE EXAMPLE

In many other types of simulations, the entire phenomenon is modeled, but in MAS, by formulating each of the micro elements that make up the phenomenon, the entire macro phenomenon created by their interaction can be expressed. Considering the example of infectious disease spread, the macro phenomenon is the infection spread process itself, and the micro element corresponds to the movement of each person who is the medium of the virus. Based on the four states of the SEIR model mentioned earlier, we construct a simple infectious disease spread example by MAS.

First, prepare a field where people can move around. Here, the structure is on the grid. Then, consider that Agent (corresponding to a person here) acts on the field every time (step) (Figure 4).

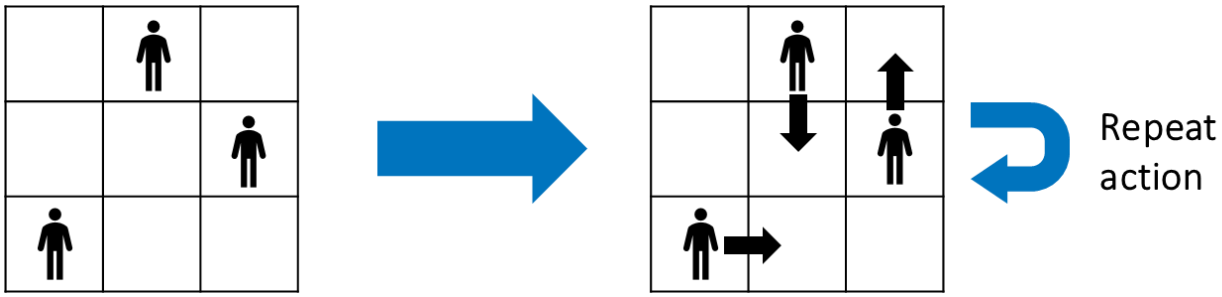


Figure 4. Multi-Agent Simulation settings in an example of infectious disease spread

By doing only this, although it is very simple, it is possible to reproduce how each person is moving independently in society. Next, the rules for infection are set as follows, for example.

It is assumed that each person (Agent) belongs to one of the four states of Susceptible, Exposed, Infectious, and Recovered as described in the previous chapter. Then, when a Susceptible person belongs to the same square as an Exposed person, exposure occurs with a certain probability, and the Susceptible person becomes an Exposed state. This time, it is assumed that infectious people will take appropriate measures such as quarantine when the onset is found, and that virus exposure will occur from Exposed, not from Infectious (Of course, depending on your assumed situation, it does not matter at all if you set that virus exposure occurs from Infectious as in the SEIR model in Chapter 1). In addition, the Exposed person and the Infectious person shift to the Infectious and Recovered states after a certain number of days, respectively.

Based on the above rules, by moving each person (Agent) and judging the status for each step, it is possible to reproduce how the infection spreads. The following (Figure 5) is an example of the progress when MAS is executed. We used the Python module *Mesa* [6] for implementing MAS.

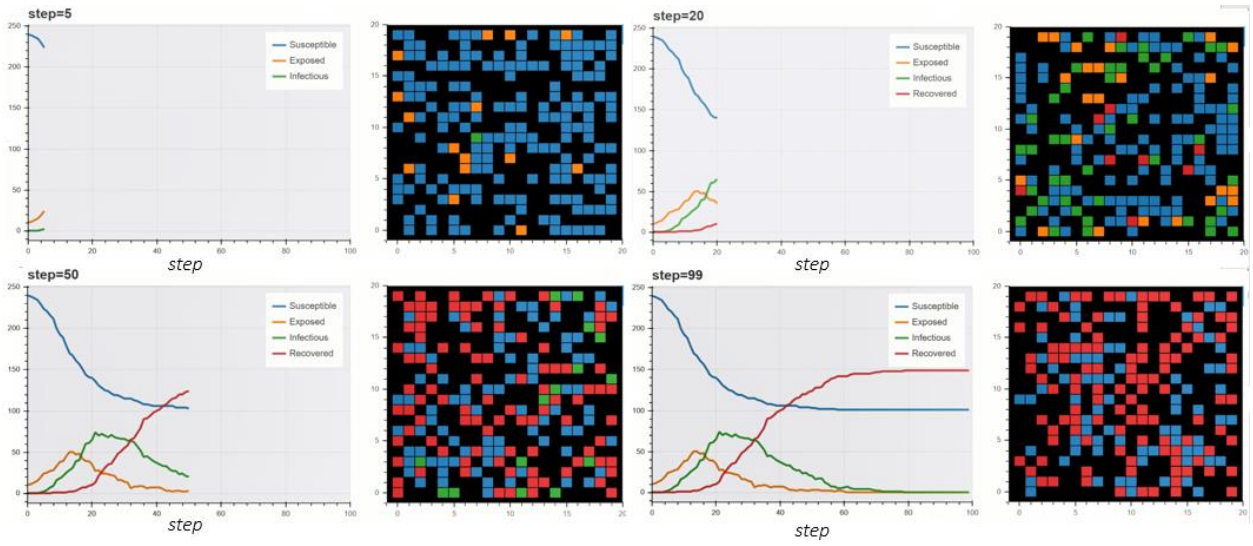


Figure 5. Infection spread progress by MAS (upper left : step=5, upper right : step=20, lower left : step=50, lower right : step=99) (blue line (square) : Susceptible, yellow line (square) : Exposed, green line (square) : Infectious, red line (square) : Recovered)

The left side of each of the four figures in Figure 5. shows the transition of the number of people in each state up to the certain step, and the right side is a grid-like society reproduced by MAS. At first (step = 5), most of the people are in the Susceptible state (blue square). Then, the infection gradually spreads, and the number of people in the Exposed state (yellow square) and the number of people in the Infectious state (green square) increases (step = 20). When the proportion of people in the Susceptible state who can be exposed and the number of people in the Exposed state who are the source of infection decrease, the infection begins to be contained (step = 50). Eventually, there are only people in the Susceptible state and those in the Recovered state (step = 99). In this way, we only decided the rules for each person's movements and state transitions, but by running the simulation according to them, we can observe the entire phenomenon of infection spread.

2.2. APPLICATION EXAMPLE

In this section, in order to show the advantages of Multi-Agent Simulation more, we consider the spread of infection in a more realistic setting than previous section. For that purpose, we construct fields such as "Home area", "Society area", "School area", "Office area", and try to reproduce the traffic between home and office, school, etc. that many people do every day in the real world (Figure 6) (Note that these settings are virtual for the introduction of MAS, and are not intended to reproduce the situation of a specific location that actually exists).

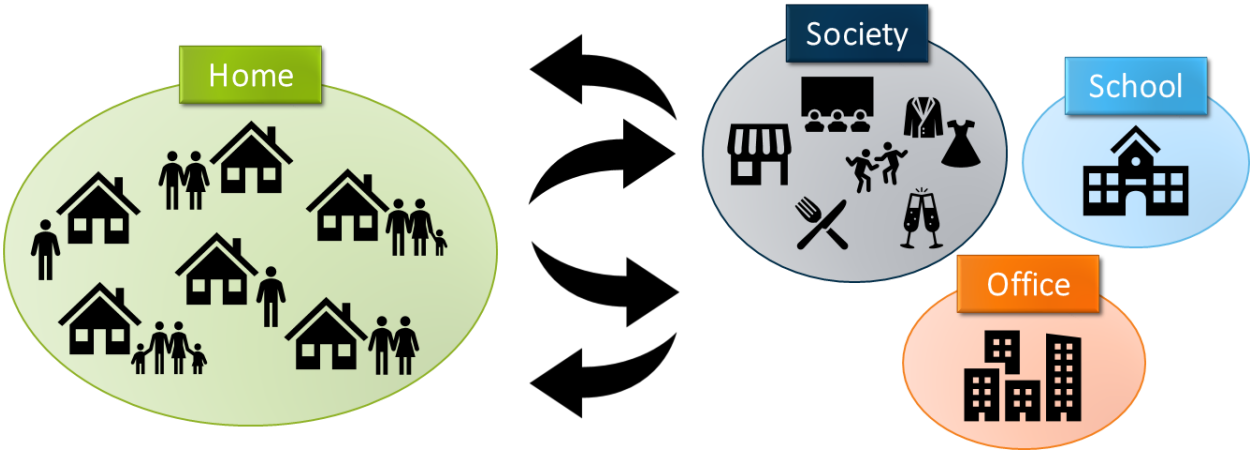


Figure 6. People's daily traffic

First, consider the same grid-like field as before, and set various areas in it as shown in the left of Figure 7. Then, each cell in the Home area (50 × 10) represents one house (so there are 500 households), and a family is assigned to each house. The number of people per household is 1 to 5 (1 person: 28%, 2 people: 32%, 3 people: 20%, 4 people: 14%, 5 people: 6%). As a result, we are thinking of a community of about 1,200 people in total. It is assumed that up to a family of two are all adults, and for a family of three or more, two are adults and the third and subsequent families are all Students. Adult agent is Worker or Housemaker, and on weekdays, Workers go back and forth between the Office area (30x40) and their homes, and Housemakers go back and forth between the Society area (20x20) and their homes (Places other than school and office, such as shopping mall and restaurant, were collectively regarded as the Society area). However, considering that there is a relatively high possibility that Worker living alone will make a detour after work, it was set to return home after moving to the Society area with a certain probability. Students, of course, go back and forth between the School area (20 x 20) and their homes. When Students and Workers are moving to the School area or Office area, the destination cell was fixed every day for each person because we think that people who usually meet are hard to

change at school and office. On holidays, everyone goes back and forth between the Society area and their home, regardless of whether they are Workers, Housemakers, or Students. In the Society area, School area, and Office area, each person moves one square or stays there for each step (1 hour), and does not move in the Home area. The time to go out is a fixed value or a value generated from random numbers. Basic parameter settings are listed in Appendix.

With the above simple settings, daily movements of people in a small community can be reproduced. Furthermore, we try to observe the spread of infectious diseases by introducing four types of states, Susceptible, Exposed, Infectious, and Recovered, as individual states. We set that the initial number of infected (Exposed) persons is 10, and all others are Susceptible. Then, in each step, when the Susceptible person and Exposed person are in the same square, it is regarded as a close contact, and the Susceptible person is exposed with a certain probability (5%) and transitions to the Exposed state. People in the Exposed state were set to develop in an average of 6 days, transition to an Infectious state, and then recover in an average of 14 days to become a Recovered state. (These values are set reasonably based on various reports, with Covid-19 in mind). While people are in the Infectious state, we assumed that they would receive treatment in the hospital, and for the sake of clarity, we set them to move to the Hospital area prepared in the grid field and to stay there.

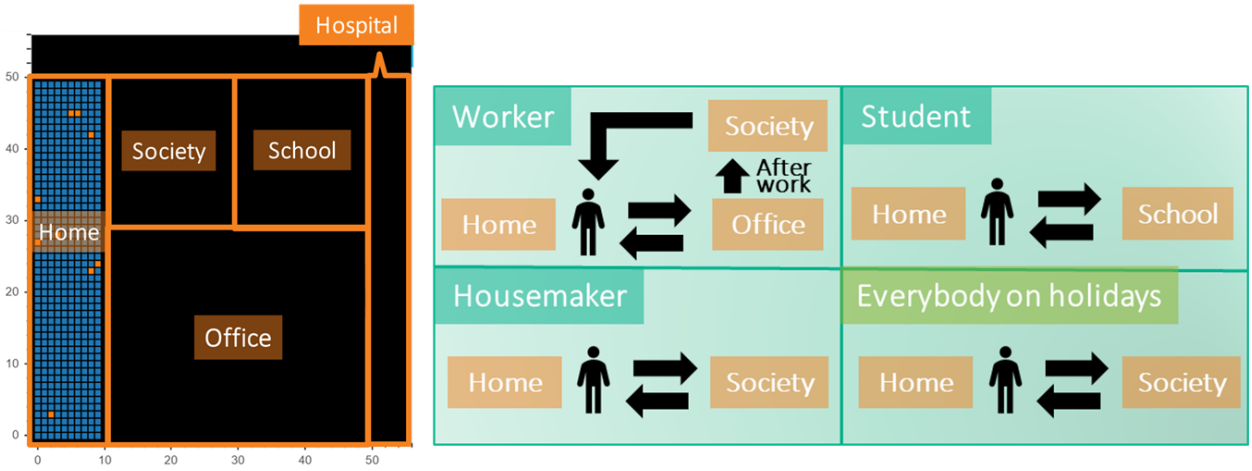


Figure 7. Field settings (left side) and behavior patterns (right side) in MAS

The following (Figure 8) is an example of the progress when MAS is executed. In the middle of the night and early in the morning, everyone stays at their own home (e.g. step = 6 : 1st day (Monday) 6 a.m.), and during the day, many go out (e.g. step = 12 : 1st day (Monday) 0 p.m.). At night, all workers return home (e.g. step = 20 : 1st day (Monday) 8 p.m.) except for some workers who stop by the Society area, and then the workers also return home (e.g. step = 24 : 2nd day (Tuesday) 0 a.m.). During the daytime on holidays, all people either stay at home or go to the Society area (e.g. step = 137 : 6th day (Saturday) 5 p.m.). As the steps are repeated, the number of Infectious people gradually increases due to contact (e.g. step = 205 : 9th day (Tuesday) 1 p.m.), and the beds (squares) in the Hospital area are filled (e.g. step = 400 : 17th day (Wednesday) 4 p.m.). By the time the pandemic has been over, most have experienced the infection once and are in a state of Recovered (e.g. step = 900 : 38th day (Wednesday) 0 p.m.).

As a result, the progress of infection spread is as shown in the left side of Figure 9. With above settings, it can be seen that most of the places where virus exposure occurs are at Home area (right side of Figure 9). Due to the setting, there is a lot of time for close contact at home, so the result is understandable. The infection may be spreading by repeating a

series of steps in which a member of a family who is exposed at home infects a member of another family at school or company, and that person infects that person's family.

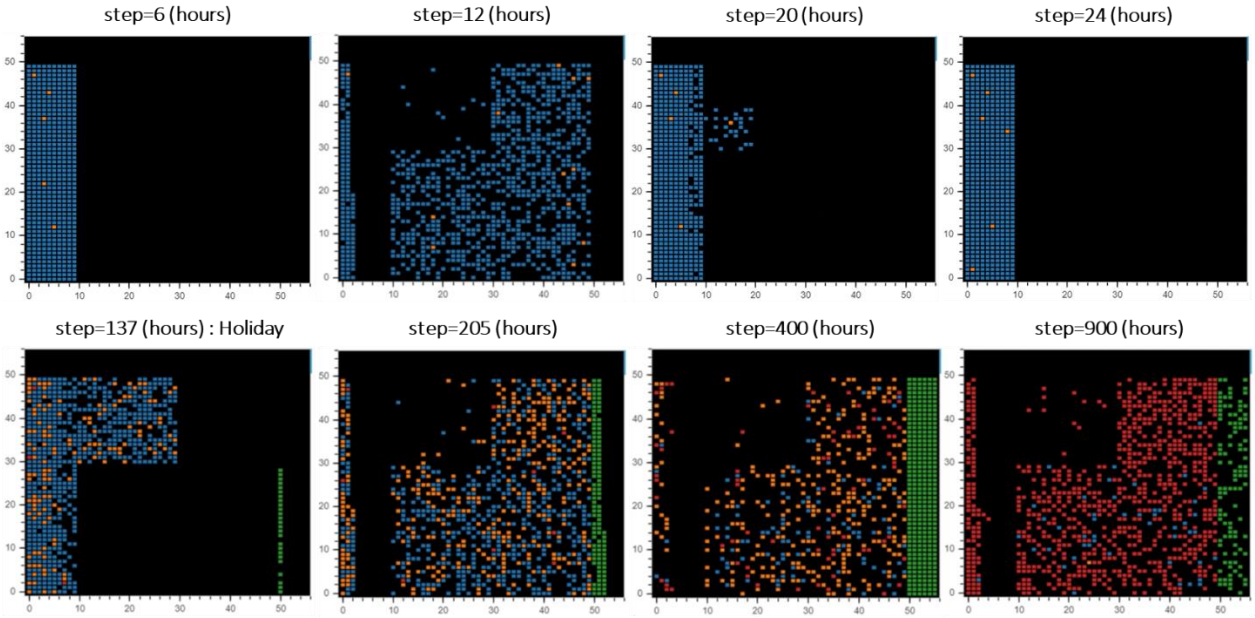


Figure 8. Progress of multi-agent simulation (blue: Susceptible, yellow: Exposed, green: Exposed, Red: Recovered)

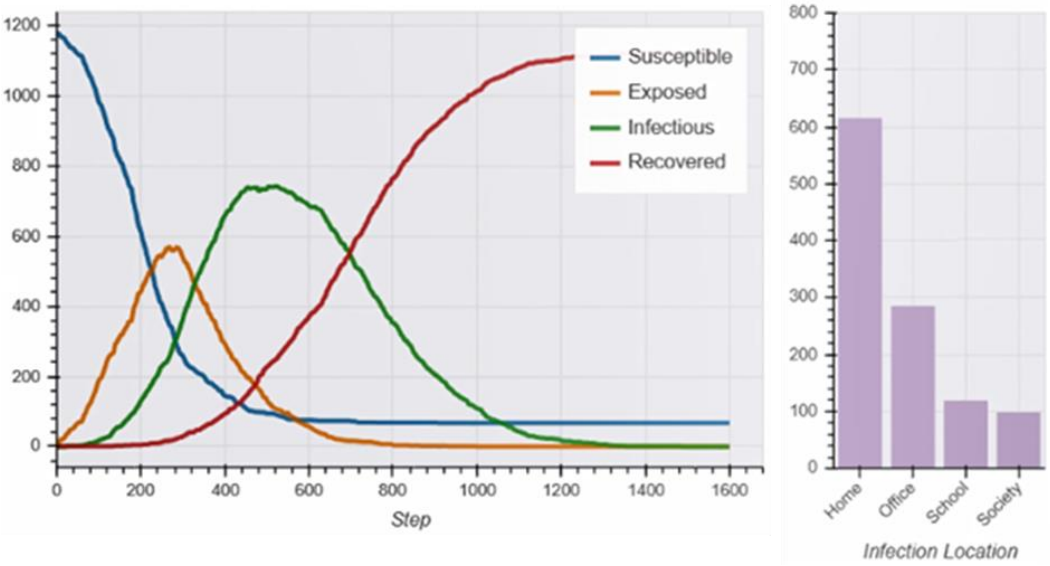


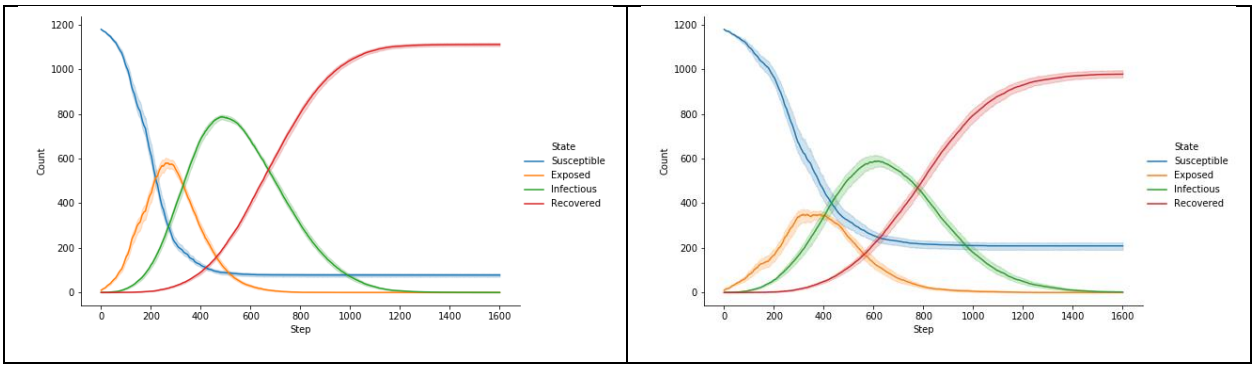
Figure 9. Progress of infection spread on MAS (left side) and number of infections by area (right side)

Looking at the left side of Figure 9, it shows that most people are finally infected, and if no measures are taken, the damage is very large. Now, let's see what happens when the countermeasures are actually taken through changing the settings of MAS. In addition to Level 0, which does not regulate anything, we consider measures (regulations) according to 5 levels. (Table 1) (Regulations for each of these Levels are set reasonably with reference to the efforts of local governments).

Table 1. Infection prevention regulations reproduced on MAS

Regulation	Wear facial coverings	Stay home after close contact with E state person uncovering	Telework (about a half Workers)	Going out restrictions	School Locked
Level 0	No	No	No	No	No
Level 1	Yes	No	No	No	No
Level 2	Yes	Yes	No	No	No
Level 3	Yes	Yes	Yes	No	No
Level 4	Yes	Yes	Yes	Yes	No
Level 5	Yes	Yes	Yes	Yes	Yes

Level 0 is the state in which no measures are taken, and when Level 1 is reached, wearing facial coverings when going out is added. This is created by reducing the infection probability during close contact when not at home on MAS (5%⇒3%). At Level 2, there is an additional restriction that people have to wait at home for some days if past close contact with the Exposed person are detected. This is set assuming the situation where a person has to wait at home until that he wasn't infected is known if someone close to him develops the disease and he is suspected to be infected. In MAS, specifically, when a person in Exposed state develops and becomes Infectious state, people who had close contact (in the same square) with him in Exposed state are set not to move from home for 14 days. At Level 3, telework is introduced in addition to Level 2 regulations. Under telework, it is assumed that about 50% of the Workers are able to work from home remotely, and that they do not move to Office area on weekdays and stay at home. Next, at Level 4, in addition to the above, it is assumed that going out restrictions are imposed. In this situation, each person's outing time is restricted (specifically, the maximum outing time of the Housemaker on weekdays is reduced from 2 hours to 1 hour, the maximum outing time of everyone on holidays is reduced from 12 hours to 3 hours, and Workers are made not to go to Society area after leaving Office area). At the highest level 5, the school are closed in addition to the regulations taken so far. Literally, on weekdays, all Students do not go to school and wait at home. The spread of infection should be greatly suppressed because schools, which are the main transmission route, disappear. In this way, in MAS, various situational changes can be flexibly incorporated and their effects can be observed (Of course, you need to pay attention to how to incorporate them. When applying MAS to a real task, it is needed to check whether the parameter values and the movements of each agent are consistent by comparing with the actual observed values in real world). The progress of infection spread observed at each regulation level is as follows (Figure 10) (Each is the transition of the number of people in each state when the corresponding regulation level is performed from beginning to end, and is shown together with the confidence interval in which the simulation was performed 10 trials).



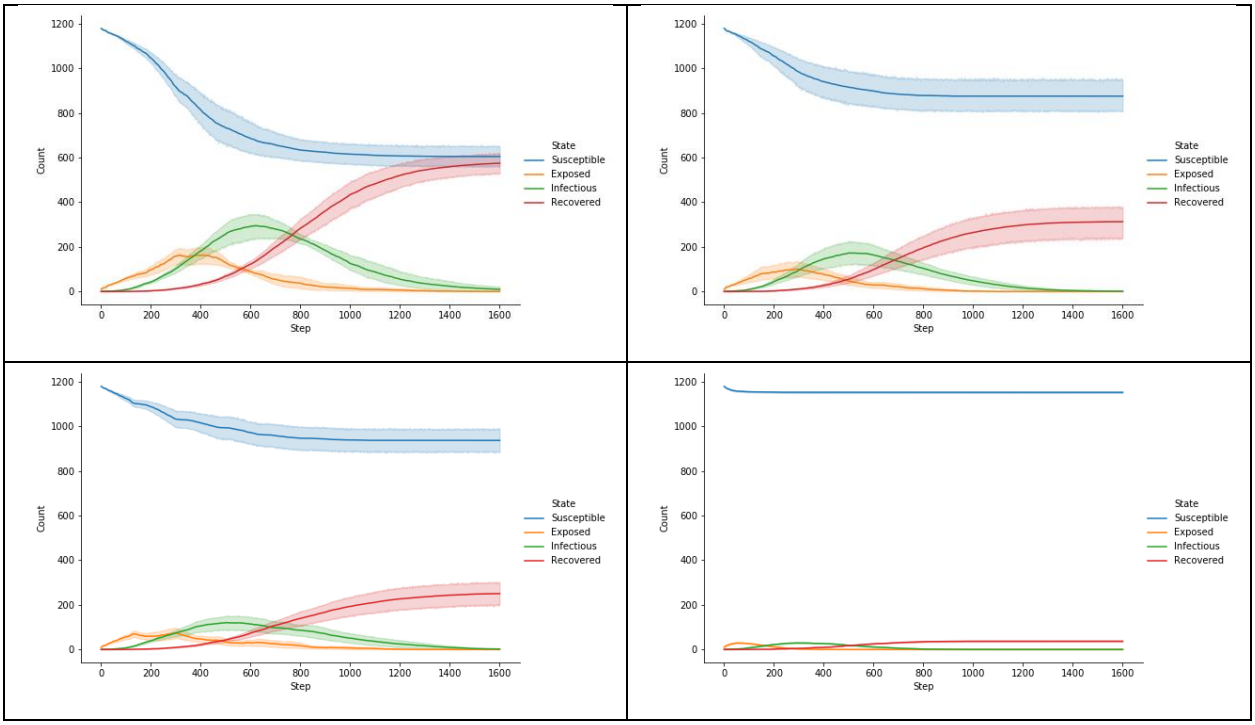


Figure 10. Infection spread progress when each regulation level is applied from the beginning to the end (upper left : Level 0, upper right : Level 1, left middle : Level 2, right middle : Level 3, lower left : Level 4, lower right : Level 5)

Looking at the spread of infection at Level 1, the maximum number of Infectious people is about 200 less than when no measures are taken (Level 0), so it is suggested that even if everyone has a little preventive consciousness in daily contact with others, it has a large effect to some extent. At Level 2 with home waiting added here, the maximum number of Infectious people has been further reduced by about 200, and the final number of uninfected people is less than the number of people who have experienced infection (the number of Recovered people). It can be said that the spread of infection can be suppressed at a minimum by making all people have preventive awareness (wearing face masks etc.) and wait at home when there is a suspicion of infection. Even in real world, if a mechanism like a contact app prevails, it may be possible to get closer to this ideal. At Level 3, which is set to telework about half of the Workers, many people have never experienced an infection and the pandemic is over. It shows the magnitude of the effect of blocking the Office area, which is one of the main infection occurrence locations, in this MAS setting. At this level, the spread of infection is largely curbed, and even if the restrictions on going out are added at Level 4, no significant difference is observed. Infections in the Society area can be curbed by restrictions on going out, that small change may be due to the fact that the number of infections in the Society area is estimated to be lower than in other areas due to this MAS setting. At Level 5, when even the School area is blocked, most of the major infection spread routes are blocked, so it can be confirmed that the spread of infection hardly occurs and is over. From the viewpoint of preventing the spread of infection, this Level 5 is ideal, but since it has significantly reduced or stopped social functions, it is a situation that should be avoided as much as possible from an economic point of view. Therefore, it is necessary to have a well-balanced infection prevention strategy that curbs the spread of infection but does not overwhelm the economy. So, we consider to realize it by switching the regulation level over time according to the situation (although we showed the transition above when the same regulation level is adopted from the beginning to the end). In other words, in situations where the infection is likely to spread, the regulation level should be raised, and as the situation becomes calm, the regulation should be switched to a lower level. That's

exactly what is happening in the real world. The problem here is which regulation level is enforced in each infection spread situation in order to eventually realize a regulation strategy that is just right from the perspective of preventing the spread of infection and the impact on the economy. This time, we focus on the utilization of Reinforcement Learning (RL) as one of the answers to this problem. By repeating trial and error by RL on the simulation, it may be possible to learn the optimum regulation strategy. It will be described in detail in the next chapter.

3. REINFORCEMENT LEARNING

In this chapter, we describe Reinforcement Learning (RL) that can be used to search for the optimal strategy in simulation including the application in the previous example of infection spread by MAS.

3.1. REINFORCEMENT LEARNING CONCEPT

First, we explain what RL is. RL is a theoretical framework that optimizes behavior through trial and error, which is one of machine learning. However, while other machine learning methods assume that big data to be learned already exists, RL newly creates and collects data within the algorithm and learn from them. In particular, it acquires data that shows the selected action and what happened as a result of taking that action. It can be said that it mimics the process by which humans and animals adapt to the environment and acquire appropriate behavior. For example, the reason why people can become to ride a bicycle is learning a sense of how to operate it without falling down by practicing many times, and not getting the correct answer from others. In RL, the subject who acts is called the "Agent", and the object on which it works is called the "Environment". In the previous example, a person is an Agent and a bicycle is an Environment. In addition, the action that the Agent performs on the Environment is called "Action", and the element of the Environment that changes according to that Action is called "State". Then, the index of the scalar value that shows the goodness of the immediate result of the selected Action is called "Reward". In the example of a bicycle, Action can be considered as the number of revolutions to pedal and how to apply the center of gravity etc., State can be considered as speed or wheel angle etc., and Reward can be considered as whether or not it falls down (0 or 1). Assuming these components, maximization of the sum of the Reward series obtained through the selection of Action in each State is aimed at in RL. In other words, search for an Action selection rule (policy) that will receive as much reward as possible in the future.

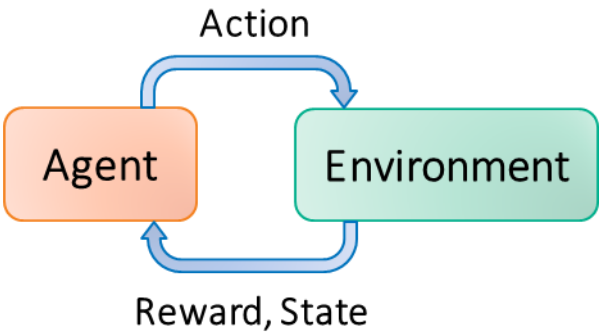


Figure 11. Reinforcement Learning framework

3.2. APPLICATION TO CASES OF INFECTIOUS DISEASE SPREAD

In this section, we show how to search for the optimal infection spread prevention strategy using RL in the environment constructed by MAS in the previous chapter. Here, the Agent in the framework of RL can be thought of as a local government that takes measures (regulations) to prevent the spread of infection, and the Environment corresponds to the world reproduced by MAS (Figure 12). Furthermore, it is necessary to set Action, Reward, and State that are exchanged between these two. Action is, of course, a regulation to prevent the spread of infectious diseases, and in particular, we consider selecting and enforcing it on a weekly basis from Level 0~Level 5 regulations prepared in the previous

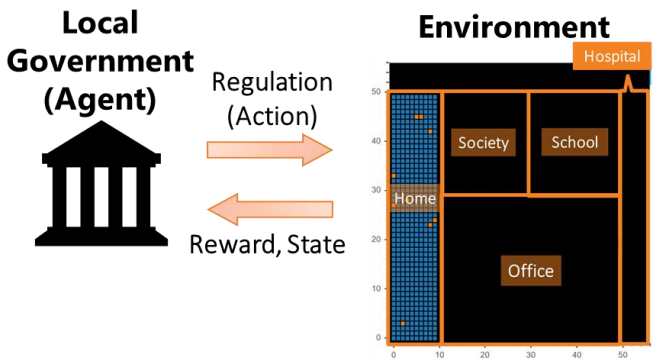


Figure 12. Reinforcement Learning framework in the case of infectious disease spread

section. Then, suppose that the number of people in each state (Susceptible, Exposed, Infectious) one week later is returned as the next State. For Reward, it is necessary to set an index that takes a large value when the spread of infection is prevented without significant economic impact (high regulation levels). Therefore, a threshold for the number of beds is set to indicate the degree of permissible spread of infection. If the number of Infectious people exceeds this number, it is considered that it has exceeded the capacity of the medical system, and Reward will be reduced.

Based on the above, the Reward is set to $\text{Reward} \coloneqq -\alpha\, max\left(\frac{n_I - n_{bed}}{n_{bed}},\, 0\right) - \beta \frac{\text{Regulation}^p}{\max \text{Regulation}^p}$ $(\alpha,\, \beta,\, p > 0)$ with reference to Kompella *et al.* (2020) [4]. Here, n_I is the number of people in Infectious state, and n_{bed} is the number of beds (set to 300). Therefore, if the number of Infectious people exceeds the number of beds, the first term takes a negative value, leading to a decrease in reward. $\text{Regulation}n$ represents the regulation level ($\in [0,1,2,3,4,5]$), and the stricter the regulation, the larger the negative value of the second term. From the above, more rewards can be obtained by relaxing the regulations as much as possible within the range where the number of people in Infectious state does not greatly exceed the number of beds. For each parameter, we set $\alpha = 0.4,\ \beta = 0.2,\ p = 1.5$. For the search for the optimal Action selection rule (policy), Q-learning (Watkins and Dayan (1992) [5]), which is a typical algorithm in RL, was simply applied by discretizing the State. Under the above settings, the 10-week infection spread process was performed 5000 times on MAS, and the regulation strategy was optimized by RL. Since it requires a lot of computing resources, we used the Python client on the SAS® Viya®, which has powerful computing power.

Simple strategies A and B are prepared for comparison with the regulatory strategies learned in RL. Strategy A is to raise the regulation level by one if the number of people in the Infectious state increases compared to a week ago, and lower the regulation level by one otherwise. Strategy B is based on regulation Level 3, which is the loosest regulation in which the number of people in Infectious state did not exceed the number of beds (=300) in the result of Figure 10. So, if the number of people in Infectious state increases compared to one week ago, the regulation level is set to 3, otherwise the regulation level is lowered by one. The progress of infection spread is as follows when the regulation strategy learned in RL is implemented and when the above two strategies are implemented (The first regulation level is set to start from 1).

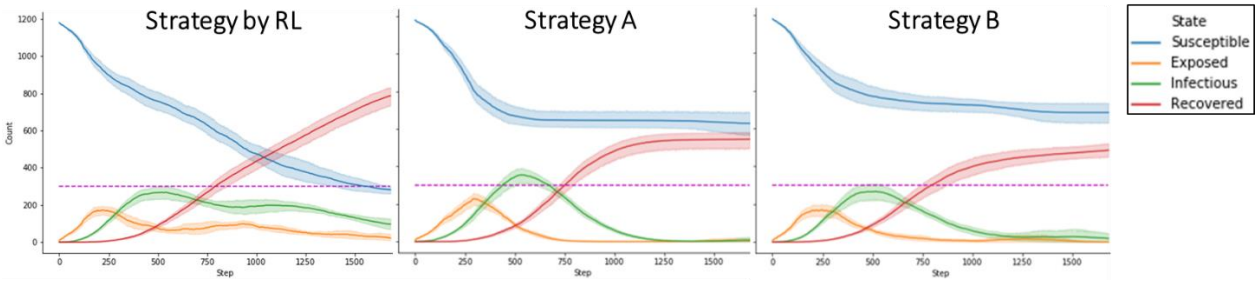


Figure 13. Progress of infection spread for each regulation strategy (dotted line : the number of beds (=300))(x-axis : step (1step=1hour), y-axis : the number of people (blue : Susceptible people, yellow : Exposed people, green : Infectious people, red : Recovered people)) (plot with confidence intervals for 10 trials)

Looking at Figure 13, the number of people in Infectious state hardly exceeded the number of beds (dotted line) in both Strategy by RL and Strategy B, while it has been exceeded it in Strategy A. Next, in order to see the impact on the economy, consider $\sum \text{Regulation}^p$ as a value representing the degree of it in each strategy (Sum is taken against 10 weeks of regulation level in one regulation strategy). The higher the regulation level, the greater this value. The average value of this value for 10 trials and the flow of typical regulation levels in each strategy are described below (Table 2).

Table 2. Economic impact of each strategy (mean of $\sum \text{Regulation}^p$) and Typical regulation level transition flow

	Strategy by RL	Strategy A	Strategy B
Economic impact	15.4	29.9	27.0
Typical regulation level transition flow	1→3→3→1→0→2→1→0→0→0	1→2→3→4→3→2→1→0→0→0	1→3→3→3→2→1→0→0→0→0

It can be seen that the strategy by RL have less impact on the economy than Strategy A and Strategy B. As you can see in Figure 13, Strategy A and Strategy B may be good in that the pandemic finishes quickly within 10 weeks. However, in terms of reducing the pressure on the medical system and also considering the economic aspect, the strategy by RL has been a good result.

As described above, by executing RL in the environment reproduced on the MAS, it is possible to generate data by oneself and promote learning. In particular, it is a powerful tool when actual data is scarce and it is difficult to be confident in empirical decision making. In this RL implementation, Q-learning was used for ease, but even for more complicated situation settings, it is possible to deal with it by using technologies such as deep Reinforcement Learning, which has been remarkably developed in recent years. Since SAS® Viya® also has the Action Set [8] for implementing deep RL, it should be possible to utilize RL in various programming environments.

CONCLUSION

This time, we introduced MAS and RL, and confirmed their usefulness through virtual examples of infectious disease spread. In MAS, it was possible to construct macro phenomena by examining micro elements, and it was possible to flexibly incorporate detailed elements into the simulation. Therefore, when it is difficult to demonstrate experiments in real world, or when characteristics of the entire phenomenon cannot be grasped and formulated, it can be one of the useful options to incorporate it into the virtual world via MAS. Of course, there is a concern that arbitrary parts such as what kind of rules should be attached to reproduce each micro element on the simulation, the range and granularity to be reproduced, and the setting of each parameter cannot be wiped out. When using it in the real task, it is necessary to pay attention to such points and consider comparison and verification with the results of actual observed data in real world. However, if you can pay attention to them, it is a powerful tool for discovering new connections and relationships between micro and macro elements that are difficult to see in reality. In addition, by involving RL here, we will fully utilize the characteristics of the simulation world that trial and error can be repeated many times, support for decision making in fields that are difficult to deal with by empirical rules can be done. In fact, in the case of the spread of infectious diseases, we were able to mechanically find out how to take a good regulation strategy from both the viewpoint of preventing the spread of infection and the economic viewpoint.

With the pandemic of Covid-19, we are now entering an era in which various environments around us are changing rapidly. Because of such a situation which it is difficult to look ahead, it is important to make the best use of simulation technology that can freely

reproduce phenomena and experiment on it in virtual space. We would like you to flexibly respond to changes in the environment by utilizing various simulation-related technologies such as MAS and RL, not only in the area of infectious diseases introduced this time, but also in various other areas.

APPENDIX

The basic parameter settings in MAS for reproducing infection spread are as follows.

parameter	**setting**
Incubation period (time to transition from Exposed to Infectious)	sample from $N(6, 3^2)$
Recovery period (time to transition from Infectious to Recovered)	sample from $N(14, 7^2)$
Infection rate (probability of transition from Susceptible to Exposed)	0.05
Distribution of the number of people in one household	(1 people, 2 people, 3 people, 4 people, 5people) = (0.28, 0.32, 0.20, 0.14, 0.06)
Number of initial Exposed people	10
Percentage of households with Housemaker (in households consisting of two or more people)	0.3
Probability of moving to Society area after leaving Office area (for people who live alone)	0.3
Hours during Housemaker's going out on weekdays (go out within 9:00 ~ 18:00)	sample from $[0,1,2]$
Hours to stay at the Society area after leaving Office area (for people who live alone)	sample from $[1,2,3]$
Hours during going out on weekend (go out within 9:00 ~ 21:00)	sample from $[0,1,2,3,4,5,6,7,8,9,10,11,12]$
Student's School area stay time	9:00~17:00
Worker's Office area stay time	9:00~18:00

($N(x, y)$ represents normal distribution with mean x and variance y)

REFERENCES

[1] Santos, G. *et al.* 2015. "Multi-agent simulation of competitive electricity markets: autonomous systems cooperation for European market modeling." *Energy Conversion and Management*, 99:387-399.

[2] Pan, X., Han, C.S., Dauber, K. and Law, K.H. 2007. "A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations." *AI and Society*, 22:113-132.

[3] Drogoul, A. and Ferber, J. 1994. "Multi-agent simulation as a tool for modeling societies: application to social differentiation in ant colonies." *Lecture Notes in Computer Science*, 830:2-23.

[4] Kompella, V. *et al.* 2020. "Reinforcement Learning for Optimization of COVID-19 Mitigation policies." *arXiv preprint arXiv:2010.10560*.

[5] Watkins, C. J. and Dayan, P. 1992. "Q-learning." *Machine learning*, vol. 8, no. 3-4, pp. 279–292.

[6] Project Mesa Team. 2016. "Mesa Overview" Accessed November 9, 2020. https://mesa.readthedocs.io/en/stable/overview.html

[7] Damien Farrell. 2020. "A simple agent based infection model with Mesa and Bokeh" Accessed November 9, 2020. https://dmnfarrell.github.io/bioinformatics/abm-mesa-python

[8] SAS Institute Inc. 2019. "Reinforcement Learning Action Set" Accessed February 25, 2021.

https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/casrlpg/cas-reinforcementlearn-TblOfActions.htm

[9] 小川 雄太郎. 2018. *つくりながら学ぶ！深層強化学習 PyTorch による実践プログラミング*. マイナビ出版.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Satoki Fujita

Shionogi & Co., Ltd.

satoki.fujita@shionogi.co.jp

Shogo Miyazawa

Shionogi & Co., Ltd.

shogo.miyazawa@shionogi.co.jp

Ryo Kiguchi

Shionogi & Co., Ltd.

ryo.kiguchi@shionogi.co.jp

Yuki Yoshida

Shionogi & Co., Ltd.

yuki.yoshida@shionogi.co.jp

Katsunari Hirano

Shionogi & Co., Ltd.

katsunari.hirano@shionogi.co.jp

Yoshitake Kitanishi

Shionogi & Co., Ltd.

yoshitake.kitanishi@shionogi.co.jp