A STUDY ON THE EFFECT OF FLOOD EXPERIENCE ON EVACUATION DECISION MAKING USING REINFORCEMENT LEARNING

RIONA MICHIGASHIRA
Department of Urban Management, Graduate School of Engineering, Kyoto University, Kyoto, Japan,
michigashira.riona.35r@st.kyoto-u.ac.jp

TOMOHARU HORI
Water Resources Research Center, Disaster Prevention Research Institute, Kyoto University, Uji, Japan,
hori.tomoharu.3w@kyoto-u.ac.jp

ABSTRACT

In the evacuation from flood caused by heavy rain in west Japan in 2018, low percentage of evacuation by residents became severe problem. It is difficult for people who has no experience of suffering damage to decide to evacuate just in case even though government issues warning earlier. In order to make those people to evacuate when it is necessary, analysis of the effect of experience will be useful. This paper tries to simulate such effect of experience using machine learning which is modeled after a process of human beings or animals learning things through gaining experiences.

Keywords: Reinforcement learning, flood experience, evacuation

1. INTRODUCTION

Japan is said to be the country most affected by climate changes in 2018 (Germanwatch, 2019). Heavy rain in west Japan was the worst flood disaster in the last 30 years, which killed more than 200 people. Even though local governments tried to issue evacuation advisory in advance, this disaster showed another problem. That is, low percentage of evacuation. In fact, in Hiroshima prefecture, which had the largest number of victims of damaged prefectures, less than twenty thousand people went to shelters while more than two million were under evacuation advisory. Many of them had stayed home and not evacuated to shelters even after they had received the information. As a result, about 70% of victims in Hiroshima were in their house when flood came. Of course not all of them might consider evacuation unnecessary but such a gap between sense of crisis of governments and that of residents also occurred in other prefectures.

It is desirable that all people under evacuation advisory go to shelter just in case before disasters happen. However, many of them have suffered little damage in most of cases fortunately. Accumulation of this lucky experience lets us judge that evacuation is not urgent. In other words, it is difficult to evacuate based on an individual judgement without actual unsuccessful experience. This is a situation that we cannot avoid when trying to fill the gap between governments and residents.

Here, it would be very useful to analyze the effect of experience on decision making for flood evacuation. Machine learning attracting a lot of attention from different areas today is suitable for this simulation because it is modeled after a process of human beings or animals learning things through gaining various experiences.

This paper is supposed to be a step on the way to reproduce and analyze the residents' sense of crisis through letting an agent experience simulated various evacuation through machine learning.

2. REINFORCEMENT LEARNING FOR EVACUATION DECISION MAKING

Machine learning can be roughly classified into three types, supervised learning, unsupervised learning and reinforcement learning. The principal difference among them is the method to learn (Odaka, 2016).

Supervised learning is the most common method in which pairs of data and answer to each case, so-called teacher data, are given. The program aims to learn the correspondence between data and answer and become able to find the answer to new cases by itself. Image recognition using convolutional neural network is one of the typical and successful examples of this type of machine learning.

Unsupervised learning is the method in which the program has some system to judge answers instead of the teacher data. The program aims to get some criteria among large amount of data by itself. This type is often used in automatic data classification.

Reinforcement learning is like a middle method between former two types. It is given reward according to the final result of a sequences of choices. It aims to find a pass which get as much reward as possible through huge number of random trials.

In the field of machine learning, someone to be trained through the program is called an agent. The objective this time is not to train the agent to evacuate at probable good timing, but to observe the effect of experiencing trial and error to protect itself from a flood. Besides, reinforcement learning is suited for evacuation simulation where we do not have enough knowledge about the environment and decision of the agent affects the result (Makino, 2016). Therefore, the method of reinforcement learning is selected for this study.

3. OUTLINE OF THE EVACUEE AGENT LEARNING FROM FLOOD EXPERIENCE

3.1 Micro Simulation Model of Flood Evacuation

Flood evacuation models have been developed considering mental attitude to risk and detailed field information for prevention and mitigation of water related hazards. In Kyoto University, the micro model simulation tools released in 2004 (Hori and Shiiba) are agent-based where people are designed to make their decisions as to whether evacuate or not and which route to take. This model enabled us to simulate the independent evacuation of people by parameterization of mental factors including danger recognition.

Based on this model, Nishikawa et al. (2019) assessed influence of sharing information about inundation. In order to find the best timing to evacuate, the evacuation simulation is performed for several scales of flood disaster changing evacuation start time by every minute. The results of these simulations are given as teacher data to learning process of the best timing to evacuate based on neural network with reference to the river water levels. The results of the various cases of evacuation simulations are used also in this study as the agent experience cases.

The target simulation area is from the midstream to the mouth of the Seri River Basin. The Seri River is an A class river, which flows from the Ryouzen Mountain into the Biwa Lake through Taga City and urban area of Hikone City, Shiga Prefecture. It has 65km^2 of total basin area and the length of the main river channel is 17km.

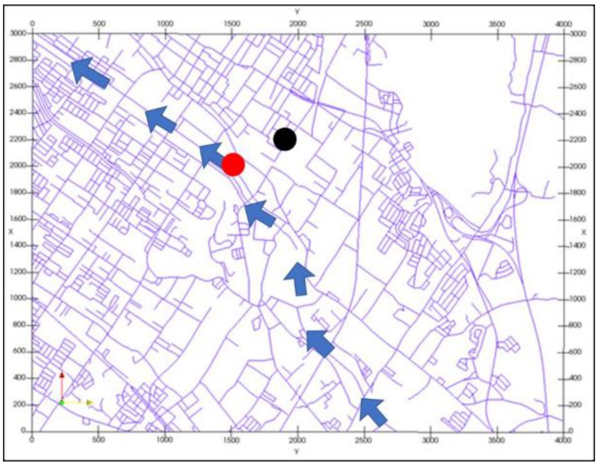


Figure 1. Positional Relation of the Target Area

In the simulation, one evacuee and his nearest shelter are considered. The positional relation is shown in Figure 1. Red circle denotes the evacuee's house, black one is the shelter, and blue arrows are position and flow of the river. Information on roads of the target area is given as arcs and nodes representing their tracks and intersections respectively. The evacuee moves on these arcs and nodes. Rainfall-Runoff-Inundation (RRI) model developed by ICHARM (International Centre for Water Hazard and Risk Management under the auspices of UNESCO) is used to simulate the dynamics of inundation. The evacuee as agent gets specific pieces of information (e.g. inundation depth) from the simulation by RRI model and leaves his house for the nearest shelter changing evacuation start time.

The evacuee is set to avoid inundated routes more than a depth of 10cm. He can know the depth of inundation in front of him within his 10m of eyesight. If current arc or node proves to be flooded, the evacuee heads back to last node he went through and searches for another route. The walking speed of the evacuee (m/s) is calculated using the depth of inundation (m);

$$v = 1.1 \times \left(1 - \frac{d}{0.7}\right). \quad (1)$$

The evacuation will fail if he is caught in inundation with the depth of more than 0.7m.

3.2 Q-learning

Q-learning is one of the typical methods of reinforcement learning. The subject to learn for the agent is called Q-value. Q-value is assigned to every selectable action and the agent chooses the next action depending on it. When some reward is given just after an action, Q-value of the action is updated. A specific calculation is as follows;

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha\big(r + \gamma \max Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)\big). \tag{2}$$

Here, s_t is the state at time t, a_t is an action selected in s_t. And $\max Q(s_{t+1}, a_{t+1})$ is the maximum Q-value of actions which is selectable at the next time, $t+1$. Three characters, r, α and γ denotes reward, learning coefficient and discount rate respectively. The value of $\max Q(s_{t+1}, a_{t+1})$ is assumed to be zero when it is the last step and there is no selectable next action.

3.3 Application of Q-learning to Evacuation Decision Making

The rough frame and flow chart of the program for this study is shown in Figure 2 and 3.

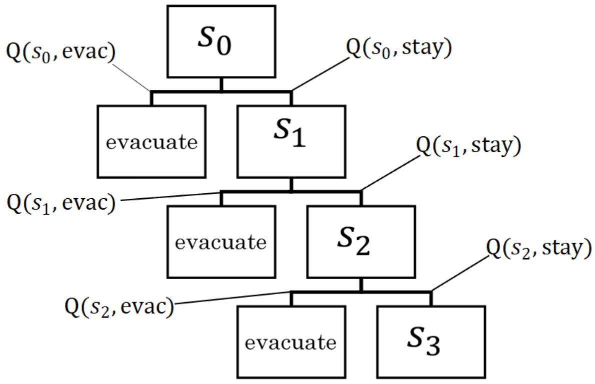


Figure 2. Frame of Decision Making

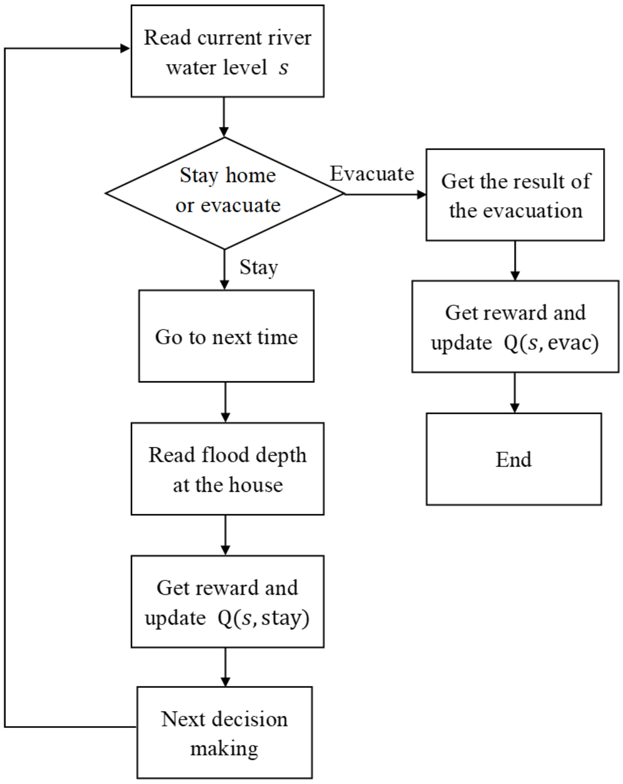


Figure 3. Flow Chart of the Program

The objective of this research is to observe the effect of experiencing trial and error in evacuations from a flood. River water level is selected as the information based on which the agent makes decision whether he should evacuate or not. The condition of inundation at his house or reachability to the shelter are considered as the criteria of judging whether the agent's decision was successful or unsuccessful. Therefore, the number of the selectable actions for each step is two, staying home or leaving for shelter. States for Q-values are river water levels at the nearest bridge to the evacuee's house in increments of 10cm. That is, there are two choices of staying and evacuating with Q-values for each river water level.

The agent receives the river water level every 10 minutes and chooses one of the two actions with a probability corresponding to the Q-value. For example, probability of choosing staying is as follows;

$$\frac{Q(s_t,\text{stay})}{Q(s_t,\text{evac})+Q(s_t,\text{stay})}. \tag{3}$$

When staying is selected, learning continues and reward is given according to the condition. Otherwise, evacuating is selected and one learning process ends by giving reward according to a success or a fail of evacuation. When the reward is zero, Q-value for the step is not updated.

This time, water level is selected rather than time as the state in order to reproduce one person's experience. The reason for this is as follows. Assume Figure 4 is a coordinate plane to express possible temporal changes of river water level during flood. There are several possible water levels at the same step of simulation according to the scales of flood events. Path 1 and path 2 are representative of temporal changes from two different scales of flood. Let an agent experience gentle flood as path 1 and consider evacuation is unnecessary. In another case, let the agent experience larger flood as path 2 and consider evacuation at s_0 necessary. The agent should estimate Q-value for staying at s_0 to be high when he learned path 1 flood experience but should estimate Q-value for evacuating at s_0 to be high when learned path 2 flood. This situation is corresponding to the cases where one person experiences a gentle flood in a certain year and also severe flood in another year. Here, Q-values at situations which never experienced are kept in initial values

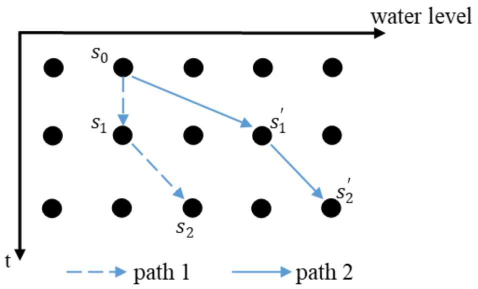


Figure 4. Possible Water Levels for Different Flood

3.4 Process for Decision Making in Flood Evacuation

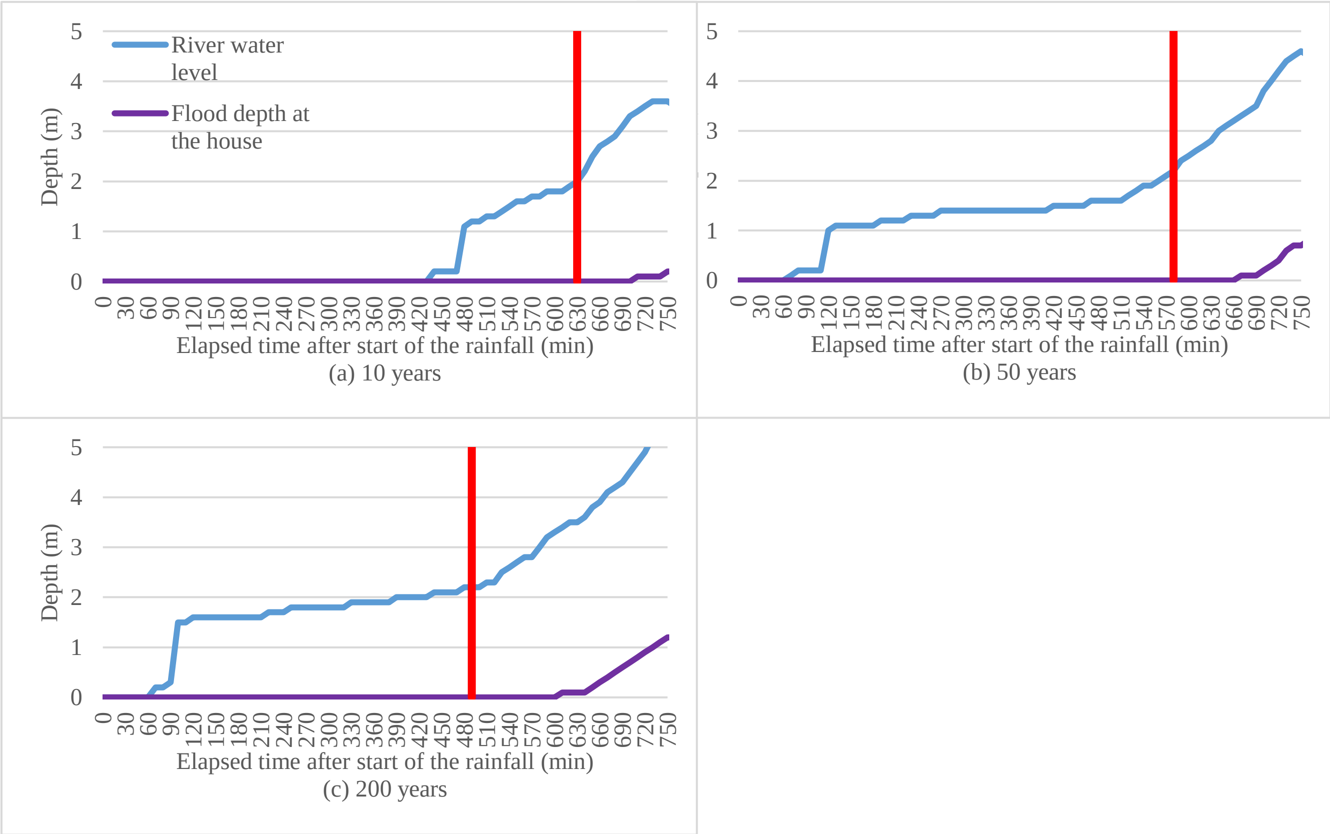


Figure 5. Flooding Conditions over Time

Table 1. Criteria for Rewarding

	Rewarding for stay	Rewarding for evacuation
Case 1	+1: flood depth at the house ≦ 50cm	0: the shelter is reachable
	−1: flood depth at the house > 50cm	−1: the shelter is unreachable
Case 2	+1: flood depth at the house ≦ 50cm	+1: the shelter is reachable
	−1: flood depth at the house > 50cm	−1: the shelter is unreachable

The simulation for several scales of floods were conducted. This paper uses the result with 10 years (the shortest), 50 years (middle) and 200 years (the longest) return period floods. Transitions of the flood condition used for this learning is shown in Figure 5. The flood depth at the house is illustrated in purple line. River water level is shown by blue line. Vertical red line in Figure 5 means the limit of time zone where the agent can reach the shelter. In 10 years return period, inundation at the house is less than 50cm and evacuation during the rainfall is rather risky. Q-learning is conducted for 750 minutes, until at the peak of inundation.

In this research, the following two cases are considered in order to reproduce the evacuee's evaluation of experience

1. No rewarding for evacuation: This case rewards the agent for staying home as long as possible. When the evacuee selected staying, he gets reward +1 if the flood depth at next step equals or less than 50cm and gets −1 if the depth is more than 50cm. When evacuating is selected, the agent gets no reward even if he can reach the shelter but gets −1 if he failed.
2. Rewarding for evacuation: Rewarding for staying is the same as in case 1 and when evacuating is selected, the agent also gets reward +1 if he can reach the shelter. Otherwise, gets −1.

Table 1 summarizes the above descriptions:

The number of trials is ten thousand. Learning coefficient is 0.1 and discount rate is 0.9. In order to prevent evacuation from concentrating around the beginning of the trial and express the tendency to stay home as long as possible, initial values of Q-value for staying and evacuating are set to 0.9 and 0.1, respectively.

4. RESULTS AND DISCUSSION

Simulation results are summarized in Figure 6 to 8. Figures on left sides show Q-values of two decisions corresponding to the river water level. Figures on right sides show the calibrated Q-values for each decision corresponding to the river water levels for each elapsed time. The upper half is for Case 1 and the lower half is Case 2. Vertical red lines are the limits to evacuate, Q-values for staying are illustrated by green lines and that for evacuating are illustrated by orange lines. Values for some river water levels are kept in their initial values because they did not appear during the simulation and not updated.

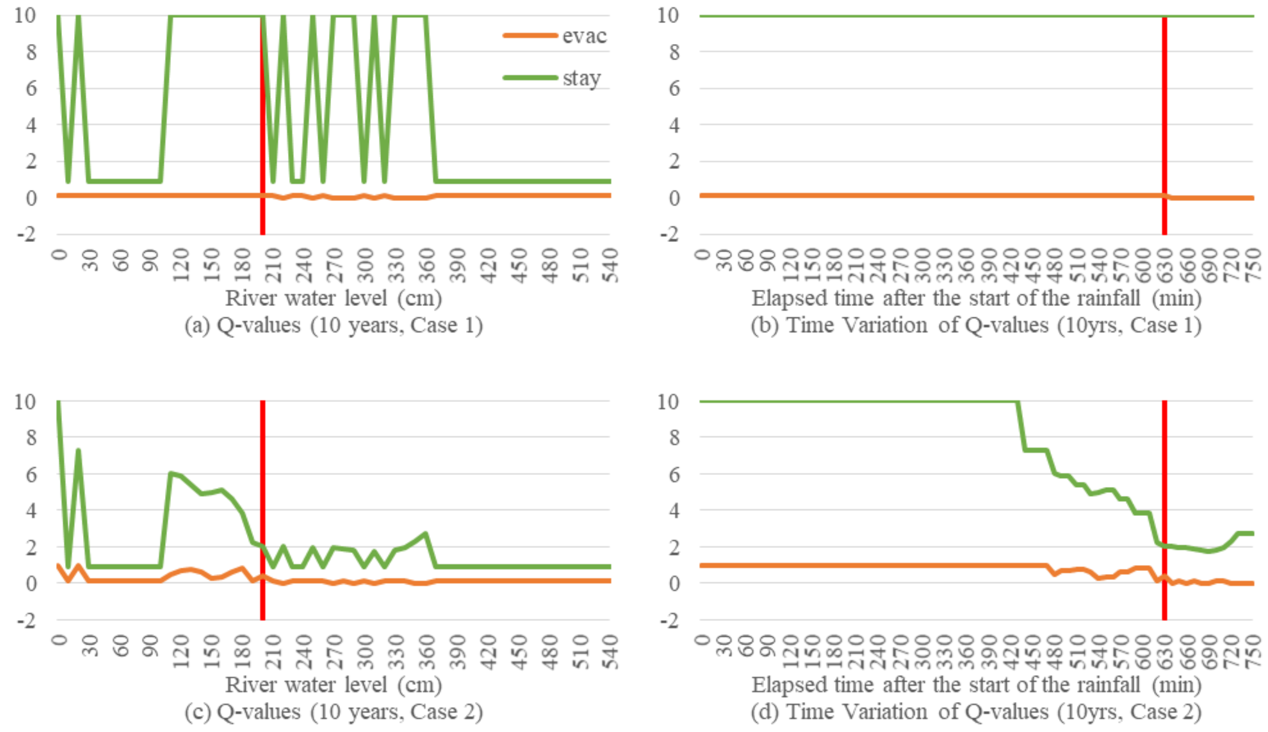


Figure 6. Results of 10 Years Return Period Flood

4.1 Flood with 10 years return period

By the structural characteristic of calculation, the maximum value of Q-value to stay is 10 and to evacuate is 1, respectively. Q-values for stay should be focused on here because the variation of the value is gradual and easy to discuss.

In 10 years return period, the depth of inundation at his house does not exceed 50cm. The route to the shelter, however, is inundated and becomes unavailable 640 minutes (the corresponding river water level is 200cm at the moment) after the start even though the shelter itself is not inundated.

In Figure 6(a) and 6(b), where Q-values for stay is almost fixed to the limit throughout the time, indicates that staying home is sufficiently safe for the agent when the foundation of the house is higher than 50cm.

On the other hand, in Figure 6 (c) and 6(d), Q-values for staying begins to fall as soon as the river water level starts to increase, almost three hours before the limit to evacuate (440 minutes) and keep decreasing sharply to much lower than Case 1 though staying is still safer than evacuation. It shows that staying home seems to be uneasy for the agent who evaluates leaving house as well. In this case, positive rewards for evacuation lead the agent to unnecessary evacuation from small floods.

4.2 Flood with 50 years return period

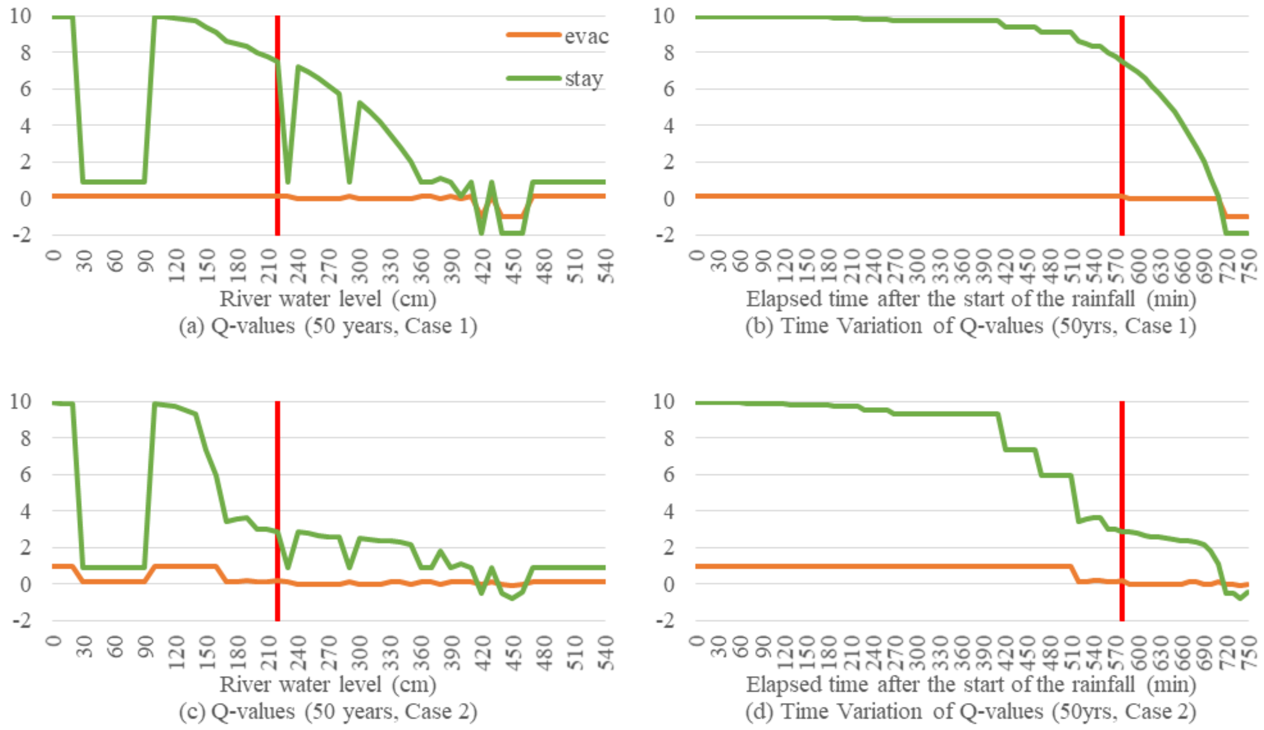
In this flood, the evacuee have to evacuate by 580 minutes (220cm) after the start. Both in Figure 7(b) and 7(d), Q-values for staying remains almost flat near 10 until 410 minutes (140cm) after the start of the rainfall. It declines from 420 minutes (150cm) to 710 minutes (400cm) after the start although it is still larger than the initial value 0.9. It becomes negative from 720 minutes (420cm) after the start, when the inundation at the house exceeds 50cm in next moment.

In Figure 7(a) and 7(b), Q-value for staying decreases more gradually than 7(c) and 7(d). Q-values for staying is quite high in Case 1 (Figure 7b) than Case 2 (Figure 7d). Therefore, there is high possibility that no rewarding for evacuation drives the agent to failure of proper evacuation decision making in this case.

4.3 Flood with 200 years return period

This flood needs evacuation by 490 minutes (220cm) after the start. Its tendency is similar to 50 years. Figure 8(b) and (d) Q-values for staying remains almost 10 until 320 minutes (180cm) and decreases to negative at 670 minutes (410cm), just before the inundation depth reaches 50cm.

What is interesting is that Q-values for staying are higher than those of 50 years return period. Especially, Figure 8(d) shows Q-values for staying at the evacuation limit is highest (almost 6) of three return periods in Case 2.



(a) Q-values (50 years, Case 1)
(b) Time Variation of Q-values (50yrs, Case 1)
(c) Q-values (50 years, Case 2)
(d) Time Variation of Q-values (50yrs, Case 2)

Figure 7. Results of 50 Years Return Period Flood

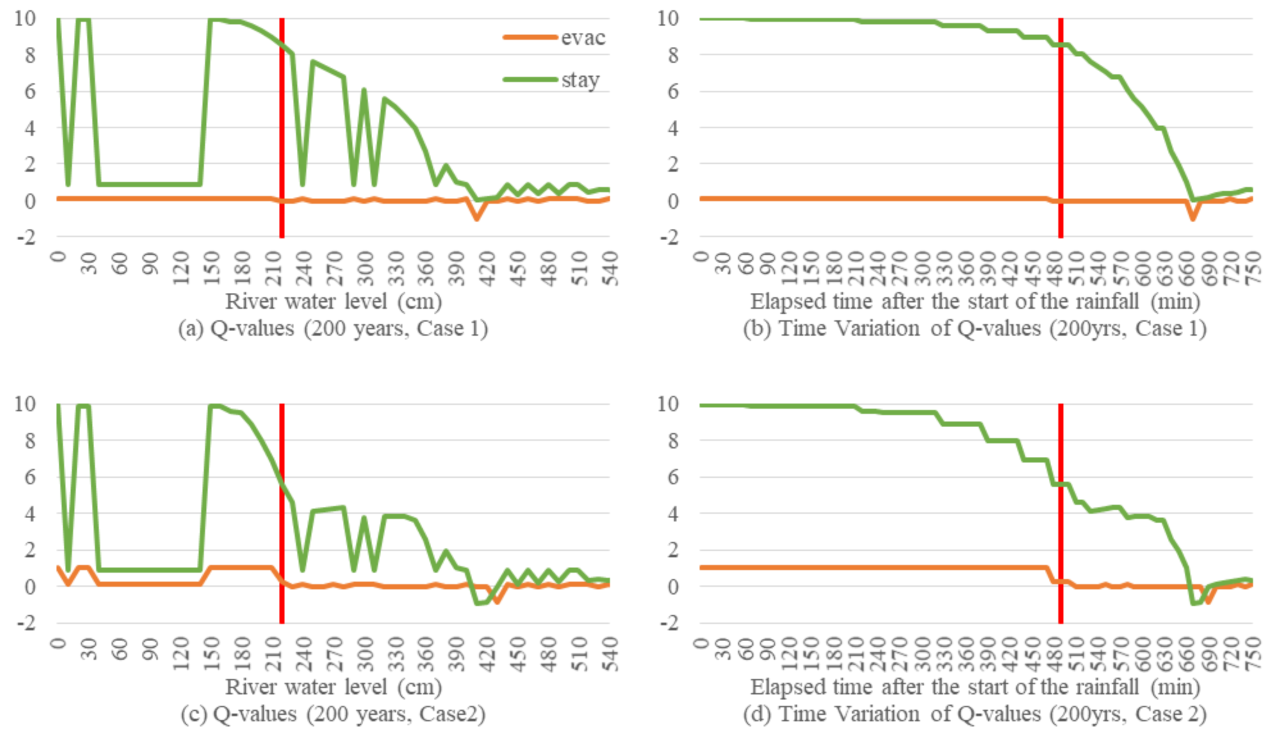


Figure 8. Results of 200 Years Return Period Flood

4.4 Discussion

From the view point of reproduction of experience, it can be said that decreasing Q-values for staying shows the effect of failure experiences in each case. Q-values for staying go below 8 around 420 minutes after the start of the rainfall in every return period of Case 2 even though they take the minimum values at different timing. This possibly indicates that a certain border of dangerous condition is estimated despite the difference of rewarding criteria or scale of the flood. Setting some threshold value will enable an agent to decide evacuation timing which takes him to the shelter safely.

However, the result of 10 years return period should be marked. Q-values for stay is low compared to larger scale floods, which suggests the agent possibly hesitates to stay home although evacuation is rather dangerous.

5. CONCLUSIONS

For a lack of real experience, there are quite a few people who will not evacuate despite the warning from the government. Today is a time when it becomes more and more important to protect your life by yourself, thus analysis of such experience using machine learning will be useful. In this paper, Q-learning is used as a tool to reproduce the effect of experience on residents' mind. Q-values are assigned to each river water levels and the machine learning is conducted with two different rewarding criteria.

Q-values for staying home decreases towards the timing to face flood and the effect of experience is expressed to a certain degree. However, Q-values for evacuating show only time zone where evacuation is reachable to the shelter. It is clear that what ideal situation to stay is like but it is still controversial problem that what to evacuate.

As future problems, there are review of rewarding criteria especially for evacuation and improvement of the system to evaluate Q-values based on more complex situations. Selection and conditions of evacuation route can effect on Q-values for evacuating. Also in situation of this research, flood reaches the house after evacuation to the shelter becomes no longer possible. If arrival of flood and deadline to evacuate come at the same time, evaluation of Q-values by an agent might be different.

REFERENCES

Eckstein, D., Künzel, V., Schäfer, L. and Winges, M. (2019). Global Climate Risk Index 2020: Who Suffers Most from Extreme Weather Events? Weather-Related Loss Events in 2018 and 1999 to 2018. Germanwatch, Germany.

Hori, T. and Shiiba M. (2004). Micro Model Simulation Tools for Performance-based Design of a Flood Risk Management System. *Journal of Natural Disaster Science*. Volume 26, Number 2, pp73-80.

Makino, T. (2016). Reinforcement Learning from Now on: 1.1 what is reinforcement learning? Morikita Publishing Co., Ltd., Japan, pp. 2–13 (in Japanese).

Nishikawa, S. and Hori, T. (2019). Effect of Information Sharing about Pathway Defect on Flood Evacuation. The Proceedings of the 8th International Conference on Water Resources and Environment Research. Nanjing, China, pp226-228. .
Odaka, T. (2016). Machine Learning and Deep Learning –Simulations by C Languages-. Ohmusha, Ltd., Japan, pp14. (In Japanese)