An analysis of optimal restoration process of Road network with Deep Q-Learning – in case of western japan flooding –

Soohyun Joo, Yoshiki Ogawa and Yoshihide Sekimoto

Abstract: Record-break flooding occurred in western Japan. Almost every road was closed owing to this disaster. Although administrative tried to recover citizen's life to normal states as soon as possible, it took a year to life all restriction on expressways [1]. In this paper, we proposed optimal decision-making considered human mobility. In addition to this, we adopt single-agent DQN to figure out optimal decision having limited information.

Keywords: Western Japan Flooding, GPS signals, Reinforcement Learning, Optimal

1. Introduction

In 2018, heavy rain fell from June 28 to July 7, in western Japan. Totally, 582 sections of the nation's main roads were closed. Government set up recovery plan to return damaged road to normal states as soon as possible. However, citizens had difficulty on daily life because reconstruction of roads that are familiar with their life wan not progressing until a week after flood occurred. We have a critical mind about present recovery work plan. In this paper, we proposed optimal decision-making considered people movement patterns. Our primary goal is to return delayed time affected by disaster to the original time. We figured out an optimal course of action administrative took using deep Q-learning.

2. Decision-Making System

We utilized GPS smartphone data from Agoop to figured out mobility patterns, and estimated Origin-Destination pairs using 1 km mesh. Because we assumed that transportation people use is vehicles such as car, bus. We identified that some origin-destination pairs enter damaged road sections with same node as its starting point. For this reason, reconstruction work in our model is assumed to be performed on specific section ODs passed by.

Soohyun Joo
University of Tokyo
Soo1992@iis.u-tokyo.ac.jp

2.1. Architecture of Proposed System

The architecture of proposed system is illustrated in Figure 1.

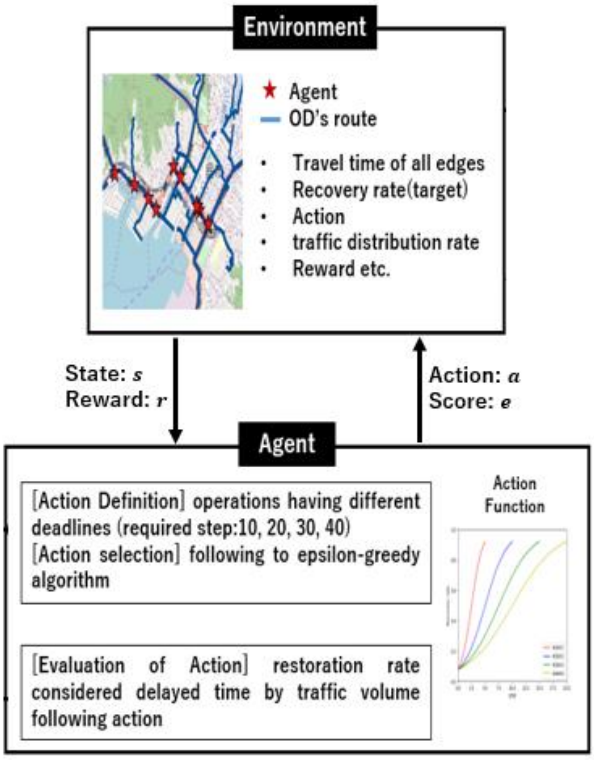

Figure 1. The architecture of proposed system

The damaged node, agent, receives the current state from environment and determines an action having the maximum its total reward. Then, the agent takes the action and the environment gives it reward and the next states corresponding to the action [2].

2.2. Assumption of Model

Actual restoration is based on the particular section of a road that has been damaged. In other words, all nodes and edges in corresponding section would have the same recovery speed. There are four assumptions in our model. First, we specified starting points as one agent. Second, the damaged section each OD pass might

recovery at the same speed as starting node's recovery. Action the start point decided affect restoration's progress. Third, we regulated work progress according to agent's damage degree. Fourth, we adopt single-agent Deep Q-Network. The single-agent framework is designed to solve stationary environments [3]. In other words, we do not consider cooperation between one agent and other agents situated near one damaged node, and assumed that agent constantly performs the action that is believe to return delayed time of all ODs passing this node to normal state as soon as possible using epsilon-greedy algorithm.

2.3. Actions

S shaped curve is one of typical theory representing cumulative progress of construction project. Actions in our model are assumed to be in accordance with sigmoid function owing to restoration is one of construction works. Thus, we identified five types of actions: 1) do nothing, 2) action requiring 10 times operation, 3) 20 times operation, 4) 30 times operations, and 5) 40 times operations until operation's progress is to be 1.

2.4. Work Progress Adjustment following Damaged Level

Even if the same restoration work was done, it takes more time to recovery the severe damaged section than others. The strength of restoration work is also different depending on the degree of hazard and operation progress. Accordingly, we calculated work progress(P_t) at time t using increment of each action's sigmoid function.

Mathematically, this equation is:

$$P_t = P_{t-1} + \Delta y * \mu$$

We adjusted the progress in accordance with damaged degree each agent had: the parameter(μ) of agent with level 3 is 0.7, that of agent with level 2 is 0.85, and other is 1.

Agent is given preference to selecting action through the reward. The basic principle is that agent with high damaged level tends to choose the action of great strength of recovery; if agent having damaged level with 3 selects action that requires 10 steps for completion, the agent receives reward with +2.

3. Experiment

We assumed that all ODs pairs would be recovered their normal state when the agent achieve restoration rate more than 0.93. The damaged node learns reconstruction action scenario to get recovery rate of 0.93 as soon as possible.

3.1. Result

In this analysis, the number of agents is 189, and the total number of ODs is 2,790. We figured out that agent takes 22 steps to get restoration rate more than 0.93 on average. Figure 2 illustrated required steps of all ODs' shortest route until achieving restoration rate more than 0.93. As Figure 2 shown, agents in Okayama Prefecture take shorter steps to meet their goal than agents in Hiroshima. The majority of road section in Okayama Prefecture have danger level with 3. Therefore, damaged nodes in this region might tend to select action with high growth rate of recovery work progress.

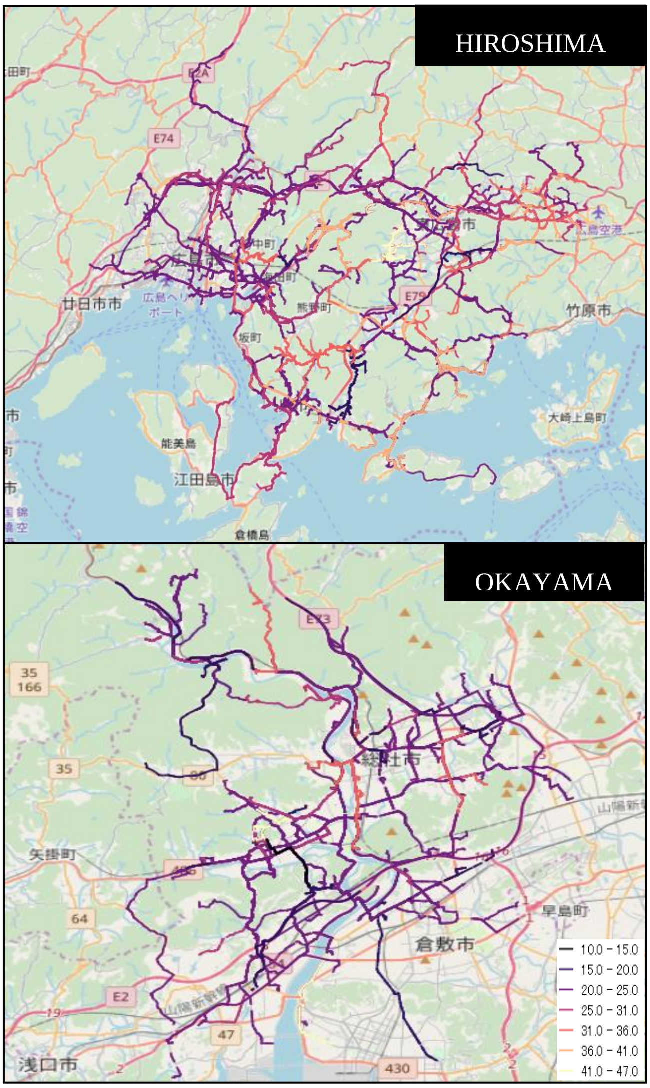

Figure 2. Required steps for completion of recovery work

Agent in DQN has a tendency to doing behavior to get the sum of high reward. Agents with high danger level are likely to choose the behavior in which the reconstruction ends with a small number of steps. In other words, the shortest route of ODs passing by an agent with high damaged level required shorter steps that other routes. Based on these characteristic, we identified that residential area has severe damage owing to this disaster.

We checked the restoration rate of every ODs' shortest route at each step. In case of Hiroshima Prefecture, almost every road section's restoration rate is more than 0.876 in 19th step. So, we focused on the aspect of the shortest route in earlier time step. Figure 3 shows restoration rate of all ODs' shortest route in 7 steps, and index on right-bottom side means 7 groups following the value of recovery rate.

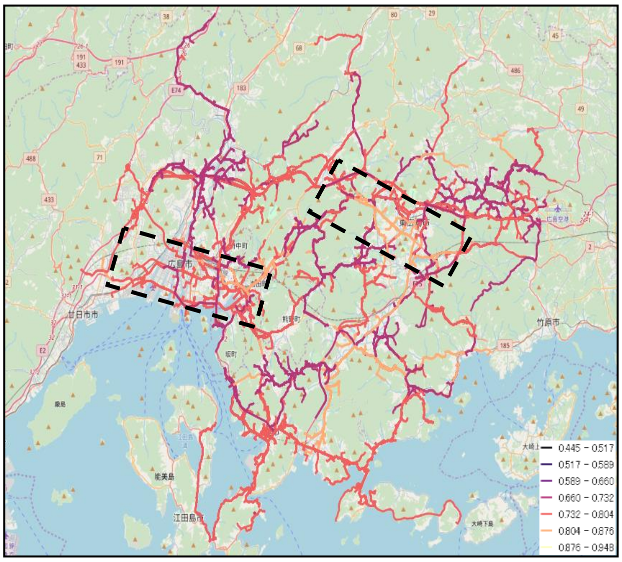

Figure 3. Restoration rate of Road sections in earlier step

We focus on two road node (agent) having same rode name, San'you express. In other words, these agents have same danger level, thus are assumed to have similar tendency related to selecting action. San'you express is located across Hiroshima city and Higashi-Hiroshima city. The left box is Hiroshima city area and the right box identified Higashi-Hiroshima city areas. As Figure 3 shown, Higashi-Hiroshima's recovery speed is faster than Hiroshima's. We believe that the agent located in Hiroshima city has more traffic volume than in Higashi-Hiroshima, thus those people would be more affected by the traffic jam as recovery work progressed. Therefore, the recovery in Hiroshima is slower than that in Higashi-Hiroshima.

3.2. Evaluation of Model

We evaluated the network, after 450 training episodes. We selected the weight to successfully get restoration rate with 0.93.

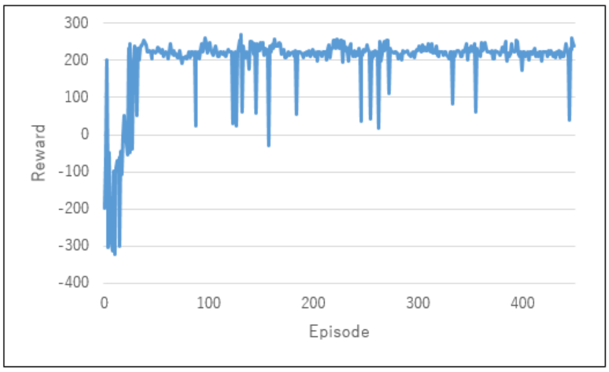

Figure 4. Reward in total training episodes

We evaluated the weight of the 250th episode, which had the highest average score. Figure 5 shows the result of evaluation with the 250th weight. Figure 5 shows restoration rate of one agent selected randomly. The average restoration was approximately 0.93 and is almost same the value of our target. In all of the episodes, the agent successfully decided a series of actions to return changed OD's state to original state as soon as possible.

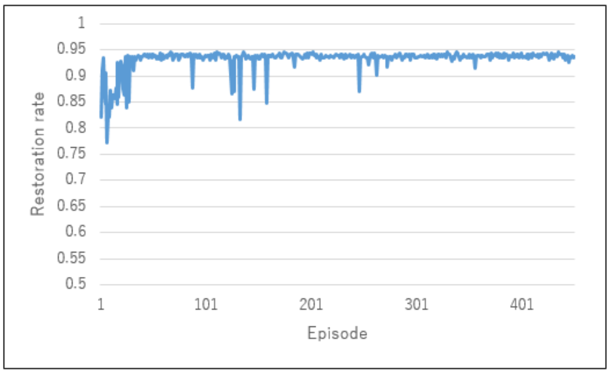

Figure 5. Restoration rate in evaluation phase

4. Conclusion

This paper presented a decision-making system about road reconstruction work using deep reinforcement learning. To proposed unique system considered mobility patterns, we utilized digital road map to reproduce real road network, Origin-Destination pairs estimated from GPS smartphone signals. We considered delayed time occurred by traffic jam, and gave action agent decided weight factor using damaged level.

Our system has limitations in that recovery work is affected by the action of starting point, agent. Unit of actual recovery work is one road section with more than 1 km length. Also, each recovery work is affected by that of other road situated with near place, and limitation of resources. Therefore, we improve the proposed system using multi-agent reinforcement learning that consider the interaction between agents. Plus, we are going to add new method about utilizing limited resources effectively.

References

[1] Response, '2018 年 7 月豪雨による高速道路の通行規制がすべて解除に', [accessed 25.08.2019] (https://response.jp/article/2019/07/08/324205.html)

[2] HongIl An and Jae-il Jung, "Decision-Making System for Lane Change Using Deep Reinforcement Learning in Connected and Automated Driving", electronics, 2019.

[3] Neto G. "From single-agent to multi-agent reinforcement leaning: foundational concepts and methods." Technical Report, Institute of System and Robotics, 2005.