Proceedings of the 2025 SICE Festival with Annual Conference
September 9-12, 2025, Chiang Mai, Thailand

A Study on Transition Rates and Similarity for Transition Learning

Satoshi Sugikawa[1†] Naoki Kotani[2] and Yuta Muraki[2]

[1] Osaka Institute of Technology, Osaka, Japan
(E-mail: satoshi.sugikawa@oit.ac.jp)
[2] Osaka Institute of Technology, Osaka, Japan

Abstract: Reinforcement learning is a machine learning technique. Reinforcement learning requires a long learning time. Transfer learning is a method of shortening this time. However, transfer learning is difficult to apply because the destination and source must be similar. Its application is based on empirical rules. Therefore, we have previously proposed a model that can formulate them. This paper especially studies the relationship between the proposed model and the transition rate. The simulations are based on the classical maze problem. The results of the simulations showed positive correlations between metastatic rates and similarity.

Keywords: Machine learning, Reinforcement learning, Transfer learning

1. INTRODUCTION

Nowadays, ChatGPT[1] has been released and is being used in all kinds of businesses. These facts show the high expectations for artificial intelligence. Reinforcement learning [2], which can learn itself, is expected to be further developed in various fields in the future.

However, reinforcement learning requires a lot of learning time because the agent learns from scratch in a new environment by trial and error. One of them is transfer learning[3]. Transfer learning is a method of adapting and reusing knowledge previously learned in a similar task for a new task. It eliminates the need to relearn, thereby reducing learning time. However, the effectiveness of knowledge is not known until it is actually transferred and learned. Therefore, when transfer learning is used, the user must consider the relationship between the transfer destination and the transfer source, but even then, learning may not be successful and negative transfer may occur.

A similarity model[4] focusing on maxQ is proposed to formalize the adaptation criterion that can discriminate the validity of knowledge in advance. In order to further demonstrate the usefulness of the proposed model, we performed simulations on the transition rate and similarity sometime in the present study.

2. MACHINE LEARNING

2.1. Reinforcement learning

Reinforcement learning is a branch of machine learning in which an agent learns through interaction with its environment. Reinforcement learning is characterized by the fact that the agent selects actions to achieve a certain goal and receives reward signals for those actions as it learns. The optimal behavior rules are learned by repeating the following steps 1 through 4.

1. Agent observes a state
2. Decides on an action based on a policy
3. Memorizes experience of which action was taken in which state and which reward
4. Seeks a policy based on experience

Reinforcement learning is formulated as a Markov decision process, expressed as M=($\mathcal{S}$, $\mathcal{A}$,T,R) as follows.

$\mathcal{S}$: Set of states
$\mathcal{A}$: Set of actions
P : State transition represented by p($s_t = s' | s_t = s, a_t = a$)
R : $\pi(s, a) = R(s_t = s, a_t = a)$ Rules of action

In reinforcement learning, the goal is to acquire behavioral rules that maximize the expected reward.

2.2. Q-learning

In this study, we used Q learning as a reinforcement learning algorithm. In Q-learning, the agent updates the Q-value associated with the combination of states and actions, and obtains an action rule that selects the action with the maximum Q-value in each state. The formula for updating the Q-value is as follows.

$$\begin{aligned} &Q(s_t, a_t) \leftarrow Q(s_t, a_t) \\ &+\eta(R_{t+1} + \gamma \max Q_a(s_{t+1}, a) - Q(s_t, a_t)) \end{aligned} \quad (1)$$

In this study, we also used the ϵ-greedy method as the agent's action selection method. In the ϵ-greedy method, the agent acts randomly with a probability of ϵ and chooses the action with the maximum Q value with a probability of 1-ϵ.

3. TRANSFER LEARNING

Transfer learning [5] aims to reduce learning time by transferring knowledge learned at the source as prior knowledge at the destination with similar tasks. However, if there is no similarity between the source and destination, a negative transfer may occur. Therefore, the user must consider the similarity between the source and destination in advance. Therefore, it is necessary to determine the similarity between the source and destination in order to determine which knowledge is to be transferred to which destination.

In this study, since Q learning is used as the transfer method, the transfer learning adopts the value function transfer type. The Q table obtained from the source agent's learning is transferred to the destination agent.

† Satoshi Sugikawa is the presenter of this paper.

The degree of reuse of the value function is adjusted by using the transfer rate τ. The transfer rate τ is adjusted in the range of $0 < \tau < 1$.

From the above, the formula can be written as Eq.(2).

$$Q^c(s,a) = Q^t(s,a) + \tau Q^s(s,a) \tag{2}$$

4. SIMILARITY INDICATORS

The proposed indicators and their comparisons are then described. The usefulness of the proposed indicator [4] has already been shown in other papers. However, we dare to include it here to facilitate comparison of transition rates.

4.1. Environment and Rewards

The algorithm used for learning in reinforcement learning is Q-learning. There are five actions that the agent can take: up, down, left, right, and no action. The (12*12) maze was prepared as the environment. the environment and actions are (12*12*5). The reward design is +1 if the agent reaches the goal, and - 0.01 if it collides with a wall or a step elapse.

4.2. Proposed model

In this proposed model, maxQ is used to determine the similarity. The total of the differences between the mazes is calculated using maxQ of maze A, which is the source of the transition, as the weight. The sum of the differences between the mazes is calculated using maxQ of maze A as the weight, and divided by the total value of maxQ.

- s is a vectorization of the source maze
- t is a vectorization of the destination maze.

From the above, the formula can be written as Eq.(3).

$$1 - \frac{\sum_i (s_i - t_i) \max_a Q(s_i, a)}{\sum_i \max_a Q(s_i, a)} \tag{3}$$

4.3. Similarity by pixel value

The first comparative indicator is the similarity by pixel value. As the mazes used in this study have the same size, the percentage of pixel values matched between the images is determined.

4.4. Cosine similarity

The second comparative indicator is the cosine similarity between the destination and source. Convert the maze to a vector to perform the calculations. Let the walls be 1 and the corridors be 0.

- s is a vectorization of the source maze
- t is a vectorization of the destination maze.

From the above, the formula can be written as Eq.(4).

$$\frac{\sum_i s_i t_i}{\sqrt{\sum_i s_i^2} \sqrt{\sum_i t_i^2}} \tag{4}$$

5. SIMULATION

We describe three models for calculating similarity, including the proposed model. The target problem is a maze problem with a range of (12 * 12). 5 maze patterns are prepared, from Figs. 1 to 5. An example of the results of the transition learning is shown in Figs. 6 to 9. The darker red areas are the locations with higher Q-values. Maze C has no red areas because the learning could not be completed.

Table 1 Result of similarity calculation

	Maze A to B(%)	Maze A to C(%)	Maze A to D(%)	Maze A to E(%)
Pixel value	82.95	77.08	80.02	79.44
Cosine similarity	80.81	75.17	77.92	76.83
Proposed model	99.01	63.32	83.2	78.3

The environment and rewards are again described here, as described in the previous chapter. The algorithm used for reinforcement learning was Q-learning. There are five actions that the agent can take: up, down, left, right, and no action. Reward design is as follows: +1 if the agent reaches the goal, -0.01 if the agent collides with a wall, or if a step elapses.

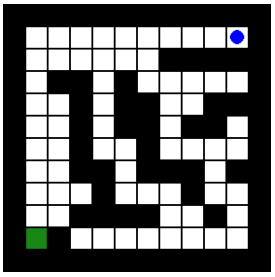

Fig. 1 Maze A

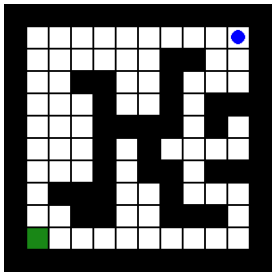

Fig. 2 Maze B

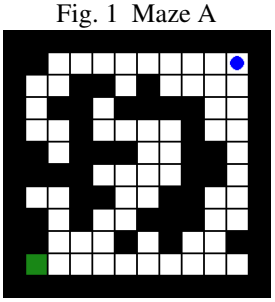

Fig. 3 Maze C

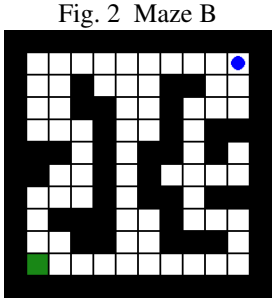

Fig. 4 Maze D

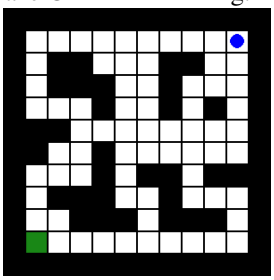

Fig. 5 Maze E

5.1. Result

The results of the similarity calculations are shown in Table 1.

The results of actually transferring the knowledge learned in are shown in Figs. 10 to 12. The red areas in the figure indicate high Q values. The left graph shows the relationship between the number of steps and episodes, and the right graph shows the relationship between rewards and episodes. However, maze C has no graph because the goal could not be reached. Four sim-

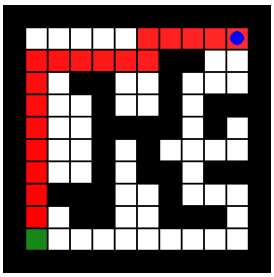

Fig. 6 Result of transfer learning for maze B.

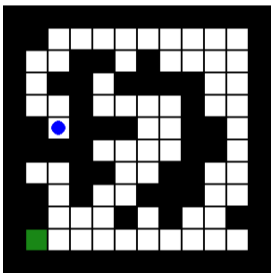

Fig. 7 Result of transfer learning for maze C.

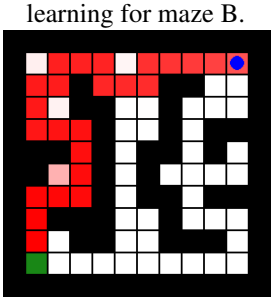

Fig. 8 Result of transfer learning for Maze D.

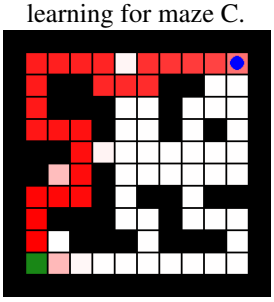

Fig. 9 Result of transfer learning for Maze E.

ulations were performed with transfer rates ranging from 0.1 to 0.9 in increments of 0.2. The red line indicates that no transfer learning was performed and that Q learning was performed.

5.2. consideration

First, the similarity is described, followed by the transition rate.

For the three similarities, maze B, maze D, maze E, and maze C have the highest similarity to maze A, in that order. The similarity by pixel value and the cosine similarity are around 80% for each maze. The proposed method shows a large difference in similarity between maze C and the other mazes where the transference fails.

For maze C, the transfer learning method learns faster than the conventional reinforcement learning method for all transfer rates. For maze D, transfer learning failed at 0.9%, was even with normal reinforcement learning at 0.7%, and generally accelerated learning at transfer rates smaller than 0.5%. The same results can be seen for maze E. These results indicate that there may be a positive correlation between the proposed similarity and the transfer rate.

From the above points, we confirmed a certain usefulness of the proposed similarity.

6. CONCLUSION

In this study, we proposed a similarity model focusing on maxQ and examined the similarity and transition rate. Five mazes were created, one as a source and the others as destinations. We compared the simulation results with the transfer learning results. The results showed that the proposed model may have a positive correlation with the transition rate. Therefore, we were able to demonstrate the usefulness of the proposed model.

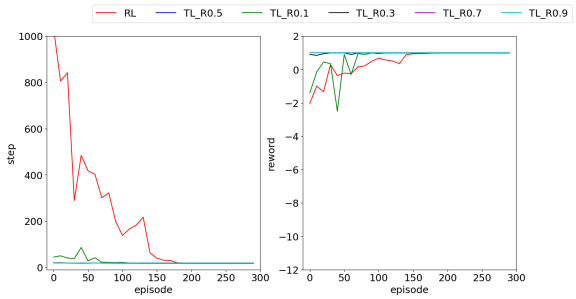

Fig. 10 Learning curve for transfer learning in maze B

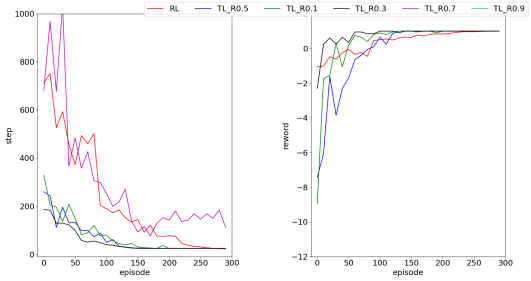

Fig. 11 Learning curve for transfer learning in maze D

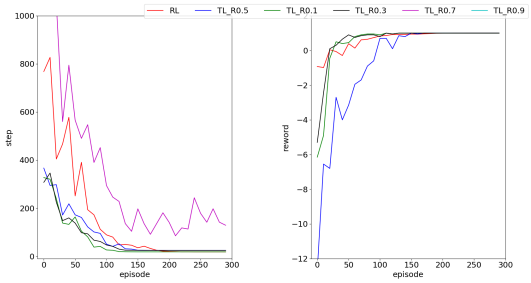

Fig. 12 Learning curve for transfer learning in maze E

Future work includes the development of effective examples for mazes of different sizes, mazes with different start/goal positions, and mazes with different types of actions.

REFERENCES

[1] OpenAI. Chatgpt(gpt-4). https://chat.openai.com, 2025.

[2] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, second edition, 2018.

[3] L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: A survey, 1996.

[4] Satoshi Sugikawa and Naoki Kotani. Similarity model for transfer learning in reinforcement learning. In *Proceedings of the SICE Festival 2024 with Annual Conference*, pp. 591–594, 2024.

[5] Matthew E. Taylor and Peter Stone. Transfer learning for reinforcement learning domains: A survey. *J. Mach. Learn. Res.*, Vol. 10, pp. 1633–1685, 2009.