ISASE 2020

Influences of pre-presented information on multi-armed bandit task

Kouhei Kudo*, Takashi Takekawa *

* *Kogakuin University of Technology & Engineering, 1-24-2 Nishi-Shinjuku, Shinjuku-ku, Tokyo 163-8677, Japan*
takekawa@cc.kogakuin.ac.jp

Abstract: Various information are taken into account on decision making. For example, we determine the transportation means or outfit of the day based on “weather forecast”. In addition, the behavior of a person also depend on the reliability of the information. In this study, we consider the case of repeating the action of selecting one from multiple choices with unknown rewards, which is well known task as multi-armed bandit, in pre-presented information. At the first time, pre-presented information is unreliable, and the effect of the information to the behavior is small. Participants can easily modify information to match the results of task if the information should be incorrect. We assume that participants may become difficult to correct the behavior due to prior information after correct information continues to be presented. It is because the reliability of information increases by repetitive correct information. Therefore, we verified how the information is corrected by the actual results and how the correction could be modified by the reliability of the pre-presented information. In particular, we analysis the participant behavior in the situation that the incorrect information is suddenly presented after the correct information is continuously presented. As a result, we found that the ratio of the incorrect choice is increased and the fixation of the behavior is delayed after the repetitive correct information.
Keywords: *cognitive bias, multi-armed bandit problem, choice behavior, incorrect information, behavior analysis*

1. INTRODUCTION

We make a variety of choices in our lives and these choices are depending on information we have. Therefore we gather the information related to the choice and use these information to estimate the resulting gain. For example, to determine whether to bring an umbrella, we receive various information such as weather, temperature and precipitation from the weather forecast.

Basically we try to select the most beneficial choice. The field of reinforcement learning predicts the reward of the next action from the history of past actions. Among them, the bandit problem is to find the best option by repeating the action of selecting one from multiple options.

Various strategies have been devised to maximize the reward, such as the "UCB strategy" [1] and "Thompson Sampling" [2] in bandit problem. But it is difficult to get the maximum reward, because we could not get enough information only from the past results. More over according to "Matching law," it has been observed that the percentage of choices matches the percentage of rewards obtained so far [3]. In other words, people may make unreasonable choices.

On the other hand, estimation is easily biased by the ambiguous information. For example, our estimation could highly depend on the pre-presented numbers even if the irrelevant numbers. The estimation about the unknown subject becomes near the anchor, which is the irrelevant number presented in advance. Such tendencies are called as anchoring effects [4]. In the case of “Weather forecast”, the information are not always true because of weather conditions.

It is considered that human behavior is biased by pre presented information. If the pre provided information is correct, we may act according to the information provided. It is also believed that the confidence in the presented information will increase. On the other hand, if the pre presented information is incorrect, there is a possibility that the information will not follow the information and will act exploratory. And the reliability of the information presented is likely to decrease. When the pre information is incorrectly presented while the correct pre information continues to be presented, people may not be able to make appropriate decisions.

Therefore, in this research, we consider the behaviors that select one from multiple choices where the reward that can be obtained is unknown. Among them, if the information on the choices is presented in advance, we will examine how the behavior has changed compared to the case where the information was not presented. It is expected that presenting the correct information increases reliability. Therefore, in the case where incorrect information is suddenly presented while continuing to

present correct information multiple times, we try to verify how a person corrects information and changes selection depend on information reliability, rewards acquired from choices.

As a result, it was found that the reliability improved when correct information was continuously pre presented. In addition, when the reliability is high, it was found that the timing of to fix the action was delayed if the incorrect information was pre presented.

2. METHODS

In this section, we describe the experiment to choice one from multiple options when pre presented information. Section 2.1 describes the detailed settings, Section 2.2 describes the pattern for pre presented information.

2.1 Experiment settings

The subject can press two buttons, "A" and "B". Figure 1 shows the website used in this experiment. When any button is pressed, "winning" or "losing" is displayed. The "Winning" is displayed as a probability, the displayed probability differs for each button, 70% or 30%. In this paper, options with a high probability of 70% are referred to as "correct option", and options with a low probability of 30% are referred to as "incorrect option". The button probabilities are different for each session and are randomly assigned. A series of flows is considered as one time, and 50 times constitute one session. There are 11 sessions. Finally, after each session, The "OK" button can be pressed. When you press the "OK" button, the information will be presented again and a new session will be started.

We will collect and analyze the data of "number of winning", "number of correct option", and "Time to choose correct option for the first time" in the behavior of one session.

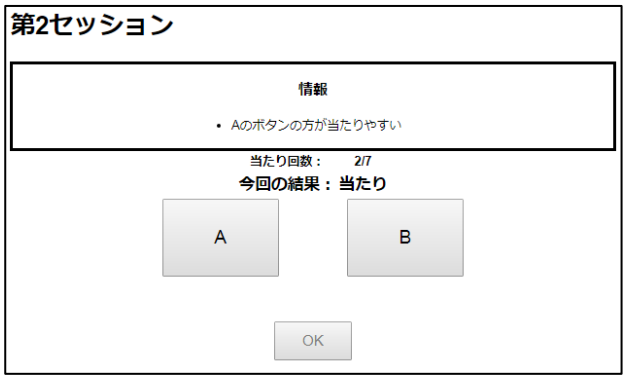

Figure 1 : Experimental website

2.2 Experiment flow

Here, we expect that the reliability will improve if correct information is presented. Therefore, in the experiment, the information presentation patterns were set to five patterns as shown in the table 1 below. For all patterns, the first session presents "no information" and the second session pre presents "correct information". Thereafter, the pre presented information differs for each pattern. At this time, "incorrect information" is always pre presented between 3 and 7 sessions. The session after the "incorrect information" is presented for the first time randomly presents the information. At this time, "correct information" is presented with a 70% probability, and "incorrect information" is presented with a 30% probability.

Table 1 : Information presentation order of each pattern (N: No <no information>, T: True <correct information>, F: False <incorrect information>)

Session	1	2	3	4	5	6	7	8～11
Pattern 1	N	T	F					
Pattern 2	N	T	T	F				
Pattern 3	N	T	T	T	F	Random (T:70%, F:30)		
Pattern 4	N	T	T	T	T	F		
Pattern 5	N	T	T	T	T	T	F	

3. RESULT

In this section, we show the experimental results. Section 3.1 shows the reliability of information, Section 3.2 shows the behavior when "incorrect information" is presented for the first time, and Section 3.3 shows how to correct the information. We recruited 100 subjects at the outsourcing service "CrowdWorks"[5]. The number of people in each pattern is 14 for pattern 1, 20 for pattern 2, 19 for pattern 3, 29 for pattern 4, and 18 for pattern 5.

3.1 Reliability increases after presenting correct information

Figure 2 shows the number of times individuals has selected a correct option in a session presented with the correct information. The horizontal axis is the session

number, and the vertical axis is the number of correct option. The data plotted is different for each session. Session 1 and 2 are data for everyone, session 3 is data for people with patterns 2, 3, 4, and 5, session 4 is data for people with patterns 3, 4, and 5, session 5 is data for people with patterns 4 and 5, and session 6 is data only for pattern 5 people.

From figure 2, it can be seen that the more times the correct information is presented, the more times the correct option is chose. If correct information continues, the reliability of the information is considered to have increased.

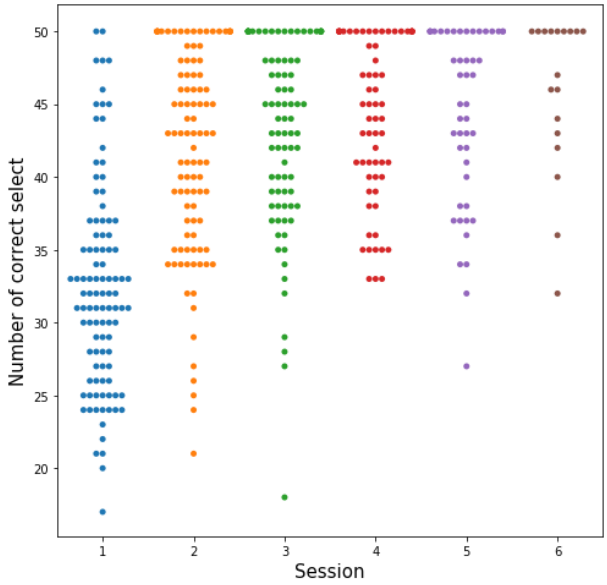

Figure 2 : Number of times you chose a correct option in each session

3.2 Behavior at the presentation of incorrect information changes depending on the reliability

Figure 3 shows an individual's choice behavior when “incorrect information” is presented. The horizontal axis indicates the number of times, and the vertical axis indicates which option is selected. In both cases, B is the correct choice, but the information presented the incorrect content that "A is easy to win".

Figure 4 shows the timing that a high probability choice is selected for the first time when “incorrect information” is presented in each pattern. The horizontal axis indicates each pattern, and the vertical axis indicates the time when the choice with the high probability is selected for the first time (the time when the information is not followed for the first time).

From figure 3, it can be seen that the individual's selection follows "wrong information" for a certain period of time. We also found that subsequent actions differed between individuals.

From figure 4, it can be seen that patterns 1 and 2 choice a high probability option at an early stage. We also found that Patterns 3, 4, and 5 take time to choose the correct option. It is considered that the higher the confidence, the longer it will take to fix the information.

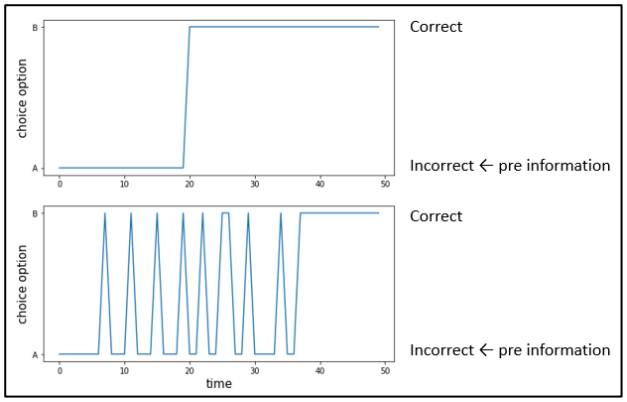

Figure 3: Example of personal behavior when "wrong information" is presented

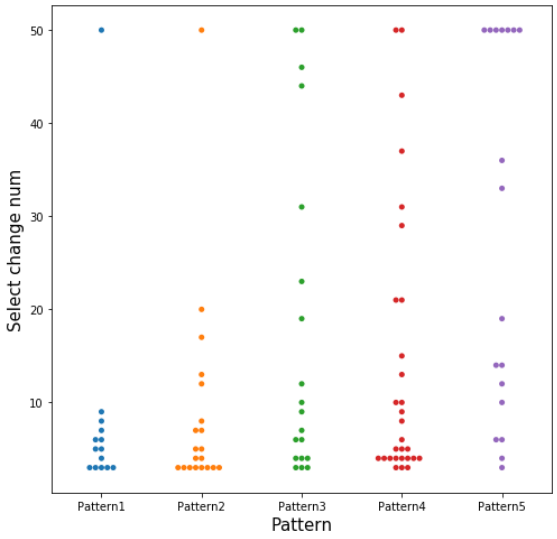

Figure 4: How often to choose the right choice when "incorrect information" is presented in each pattern

3.3 Fix of information is in the second half

Figure 5 shows the average of the selection ratio when “incorrect information” is presented. The horizontal axis indicates each pattern, and the vertical axis indicates the number of selections. Here, the bottom of the bar indicates the number of choices with a high probability, and the top of the bar indicates the number of choices with a low probability. The bar on the left indicates the ratio of the selection from the first to the 25th, and the bar on the right indicates the ratio of the selection from the 26th to the 50th.

From figure 5, except for pattern 1, the number of

choices with high probability increased in the latter half. It is considered that the information was corrected in the latter half of the choice. .

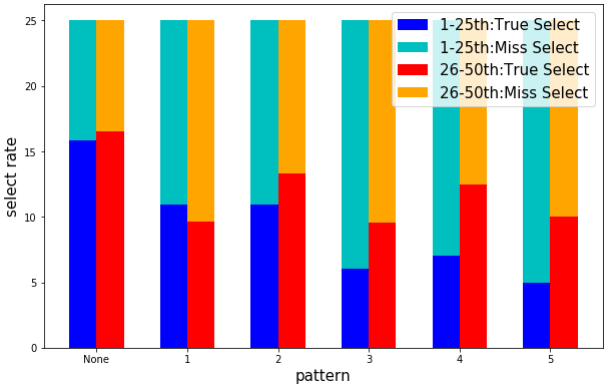

Figure 5 : Selection ratio of each pattern when incorrect information was presented

4. Conclusion

In this experiment, we verified whether the reliability of the information affected the selection. From the results, it was found that the higher the reliability, the higher the possibility of selecting an option with low probability according to "incorrect information". And the information will be modified later in the session.

This study show that there was a qualitative tendency to modify the information provided. Therefore, it is expected to be produce a quantitative model based on the relationship between the reliability of information and the behavior at the time of incorrect information in the future.

REFFERENCES

[1] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. Machine Learning, vol.47, pp.235-256, 2002

[2] D. J. Russo, B. Van-Roy, A. Kazerouni, I. Osband, and Z. Wen. A Tutorial on Thompson Sampling. Foundations and Trends in Machine Learning, vol.11, no1, pp 1-96, 2018.

[3] R. J. Herrnstein. Relative and absolute strength of response as a function of frequency of reinforcement. Journal of the experimental analysis of behavior, vol.4, pp.267-272, 1961.

[4] A. Tversky, D. Kahneman. “Judgment under Uncertainty: Heuristics and Biases,” Utilty, probability and human decision making. Springer Netherlands, vol.185, no.4157, pp.1124-1131, 1974.

[5] CrowdWorks
URL: https://crowdworks.jp/dashboard