Modeling Intrinsic Motivation in ACT-R : Focusing on the Relation Between Pattern Matching and Intellectual Curiosity

Modeling Intrinsic Motivation in ACT-R: Focusing on the Relation Between Pattern Matching and Intellectual Curiosity

Kazuma Nagashima (nagashima.kazuma.16@shizuoka.ac.jp),
Junya Morita (j-morita@inf.shizuoka.ac.jp) ,
Yugo Takeuchi (takeuchi@inf.shizuoka.ac.jp)
Department of Informatics, Graduate School of Integrated Science and Technology, Shizuoka University,
3-5-1 Johoku, Naka-ku, Hamamatsu-shi, Shizuoka-ken, 432-8011 Japan

Abstract

To be keen learners, humans need both external and internal rewards. To date, many studies on environmental learning using intrinsic motivation for artificial agents have been conducted. In this study, we aim to build a method to express curiosity in new environments via the ACT-R (Adaptive Control of Thought-Rational) cognitive architecture. This model focuses on the "production compilation" and "utility" modules, which are generic functions of ACT-R, and it regards pattern matching with the environment as a source of intellectual curiosity. We simulated a path-planning task in a maze environment using the proposed model. The model with intellectual curiosity revealed that understanding of the environment was improved through the task of searching the environment. Furthermore, we implemented the model using a standard reinforcement learning agent and compared it with the ACT-R model.

Keywords: Cognitive modeling; intrinsic motivation; ACT-R

Introduction

Humans can learn in a wide range of environments to achieve their goals using rewards generated internally and externally. To simulate such keen learning through artificial agents, the concept of intrinsic motivation, which is driven by rewards, such as self-efficacy and curiosity, has been discussed by several researchers (Manoury, Sao, & Cédric, 2019; Schmidhuber, 2010; Singh, Barto, & Chentanez, 2005).

However, these researchers have not explained the connection of intrinsic motivation with other primitive cognitive functions based on the framework of reinforcement learning. In contrast, recent studies of cognitive modeling have increasingly relied on cognitive architectures, which integrate primitive functions commonly used in various tasks (see Kotseruba & Tsotsos, 2020, as a recent review). By sharing primitive processes between different tasks, the overall structure of human cognition is defined.

Following these trends of studies, the current study aims to present one of the possible cognitive mechanisms behind intrinsic motivation based on cognitive architecture. Among several cognitive architectures, we use ACT-R (Adaptive Control of Thought-Rational; Anderson 2007). This architecture has been widely used, and considerable research has been conducted on it. Furthermore, ACT-R has a module similar to that of reinforcement learning used in conventional autonomous agents. Thus, we consider it useful to model intrinsic motivation in ACT-R by connecting the basic cognitive functions that have already been validated by various psychological experiments.

Before presenting the model, we clarify our purpose by reviewing the previous studies relating to this topic. Following this, we propose a mechanism of intrinsic motivation, which especially focuses on pattern matching between the environment and internal knowledge, assuming a correlation to human intellectual curiosity. The proposed mechanism is implemented to run simulations of a specific task. Finally, we summarize the current status and indicate future directions of research.

Related Works

This section presents two directions of studies about environmental learning: studies based on reinforcement learning and ACT-R.

Intrinsic Motivation in Reinforcement Learning

To date, several researchers have studied artificial agents with intrinsic motivation (Manoury et al., 2019; Schmidhuber, 2010; Singh et al., 2005). These studies have modeled curiosity, which is one type of intrinsic motivation, and have investigated methods to make agents search the environment widely. Such studies have primarily used statistical learning frameworks, such as reinforcement learning. Usually, agents created from reinforcement learning determine their actions based on information received from the external environment. The environment generates rewards depending on the result of their actions, and they seek to maximize the rewards it over time. Regarding this traditional framework, Sutton and Barto (1998) pointed out that the boundaries between agents and the environment are not the same as the physical boundaries between the body and the environment. Following this claim, Singh et al. (2005) proposed intrinsically motivated reinforcement learning (IMRL). In contrast to conventional reinforcement learning, in which one receives a reward directly from the external environment, IMRL fluctuates depending on the state of the internal environment and models the curiosity for an unexpected response.

In recent years, this topic has remarkably progressed with a framework of deep reinforcement learning (Burda et al., 2018; Pathak, Agrawal, Efros, & Darrell, 2017). Burda et al. (2018) examined environmental learning based solely on intrinsic rewards. The screens of games, such as Atari and Unity maze tasks, were used as input (Mnih et al., 2015), and internal rewards were generated from novel experiences for agents. As a result, agents learned a wide range of environments and improved their game scores. The authors indicated that game environments are usually designed to stimulate the users' curiosity, and the game scores increase when they find

new information in the environment.

Environmental Search and Emotion in ACT-R

The studies presented in the previous section did not examine the association between humans and models but aimed to propose learning algorithms that realize optimal searches. In contrast, ACT-R is a cognitive architecture with modules corresponding to brain regions. For example, the declarative module retains experience and knowledge, and the goal module manages states in tasks. The production rules in ACT-R are selected based on the status of such modules, and they send commands to the modules as actions (e.g., search for knowledge that meets the conditions and update the current state of the task). These rules include variables that realize flexible correspondence (pattern matching) with module states.

Concerning environmental learning, Fu and Anderson (2006) used ACT-R to solve the repeated maze task by implementing knowledge concerning direction, such as up, down, left, and right. Their model used an ACT-R function called the "utility module", which is similar to Q-Learning, a model-free reinforcement learning algorithm to optimize a policy of determining action taken under a specific situation (Watkins, 1989). Using this module, the model received a positive reward when performing an action leading to achieving the current goal and a negative reward when performing an action that not leading to achieving the current goal. By doing this, the model learned to take optimal action by increasing the number of trials.

Other research performed path planning of the maze task using not only reinforcement learning but also learning of declarative knowledge, which was implemented in ACT-R. Reitter and Lebiere (2010) employed declarative knowledge representing the structure of mazes as topological maps and presented a backtracking algorithm searching the topological maps to find a goal. Their model did not include implementations of acquiring such topological maps but assumed that they were acquired through a general mechanism of instance-based learning that uses experience to solve the current situation (Gonzalez, Lerch, & Lebiere, 2003).

Although the official ACT-R theory has so far not directly included the topic of intrinsic motivation for environmental learning, many researchers have been working on models of emotion, which relate to intrinsic motivation. Dancy, Ritter, Berry, and Klein (2015) explained the influence of emotions by combining cognitive processes of ACT-R with physiological mechanisms. van Vugt and van der Velde (2018) built a cognitive model that describes depression by the proportion of memories with positive and negative emotions. Furthermore, Juvina, Larue, and Hough (2018) constructed a model of learning emotional memories using internally generated reward functions. Each of these studies developed novel modules or functions of ACT-R to approach emotional processes. In contrast, in this research, we aim to model intrinsic motivation using the existing built-in functions of ACT-R. In particular, the current study proposes a mechanism of reward fluctuation that naturally emerges from the learning process of ACT-R instead of directly defining reward functions as a formula.

Proposed Mechanism of Intrinsic Motivation

This section presents our proposed mechanism of intrinsic motivation. The mechanism is based on the idea of connecting intellectual curiosity with pattern matching. After describing this idea, we present a general framework of intrinsic motivation by combining the existing functions of ACT-R.

Intellectual Curiosity and Pattern Matching

Following Burda et al. (2018), we focus on *curiosity* as one of the causes of intrinsic motivation. As shown in previous studies, the agents' curiosity facilitates the exploration of the game environment, and the agents' game performance improves. In the book *Theory of Fun for Game Design*, game designer Koster (2004) said that good games stimulate users' curiosity. He also mentioned that the *fun* in the game is defined as *discovering patterns* leading to continuous learning. For example, in games where the optimal solution is found from several patterns, there is nothing to be obtained from the game after finding the optimal solution, and boredom occurs.

In the current study, we focused on the *pattern-matching* mechanism as a concept analogous to the discovery of patterns by humans. Pattern matching is a primitive-purpose mechanism. For a popular example, not limited to cognitive modeling, even text searching uses pattern matching, which is expressed as regular expressions. In ACT-R, as mentioned, pattern matching is used to match production rules and module states. Figure 1 explains pattern matching in ACT-R. In this example, *Variable 1* and *Variable 2*, which are included in the then clause of the ACT-R production rule are matched with the constants (i.e., numbers such as 1 and 2) of the declarative knowledge.

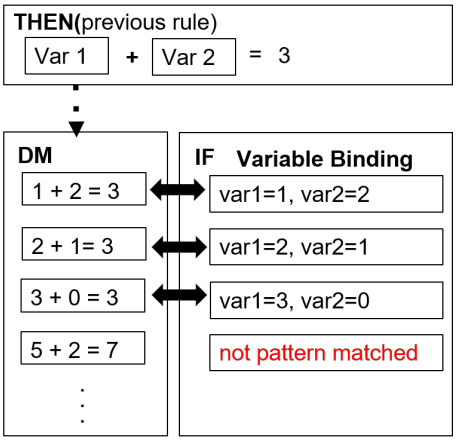

Figure 1: Example of ACT-R pattern matching. The model queries "declarative memory (DM)" with the "THEN" of the previous rule and illustrates the flow in which variables are bound by the "IF" of the next rule.

In this way, pattern matching discovers structures in the environment according to the patterns of variables embedded in the rules. Anderson (2007) claimed that a type of pattern matching dealing with relational structure is essential to achieving human-specific cognitive functions, such as cognitive flexibility, linguistic processing, metacognition, and analogical reasoning[1]. From this claim, we assume that the pattern matching of the cognitive model might lead to a model of human intellectual curiosity based on pattern discovery.

Decay of Intellectual Curiosity

To explain the role of intellectual curiosity in the primitive cognitive process, we need to consider how such a motivation decays during the process of task execution. We assume that such a decay process is the reverse of learning, namely boredom. To consider this in detail, the following subsections present a summary of the learning functions of ACT-R: the *utility* and *production compilation* modules.

Utility The ACT-R has two types of knowledge: declarative (chunks) and procedural (production rules). Each has learning mechanisms to acquire and to modulate the use of knowledge. Among these approaches, we focus on the modulation of procedural knowledge to determine whether to continue performing the task (motivated state) or to quit the task (bored state). As noted in the previous section, ACT-R has a utility module that controls a conflict resolution (selecting one of several rules that can fire (execute) in a specific situation) and updates the utilities through rewards (Fu & Anderson, 2006). Using this module, we solve a conflict between the task continuation rule and task stopping rule and assume that the number of rewards adjusting these utilities is influenced by the execution of the production compilation.

Production Compilation The production compilation module combines two successive production rules into one production rule (Taatgen & Lee, 2003). By repeatedly firing a series of rules for a certain task, the integration of rules occurs, and the number of rules that fire is reduced until the task is completed. Production rules that are the target of integration usually include variables in the conditional clauses (the IF parts). In ACT-R, the flexible nature of human thought is modeled by pattern matching of declarative knowledge with the pattern of variables described in a production rule. The integration of rules by production compilation skips such flexible pattern matching (i.e., avoiding retrievals of declarative memory). In other words, variables contained in the rules before integration are replaced by static values copied from declarative knowledge, and routine automatic task execution procedures are produced.

Figure 2 illustrates a trace of the ACT-R model that plans the path from the current position to the goal position in a maze environment, which is the task used in the simulation presented in the later section. The vertical axis indicates time, and each column indicates a module event. The trace on the left represents the initial state of the model using planning declarative knowledge to plan the path. The trace on the right is the process after the compilation when the model plans the path without retrieving declarative knowledge.

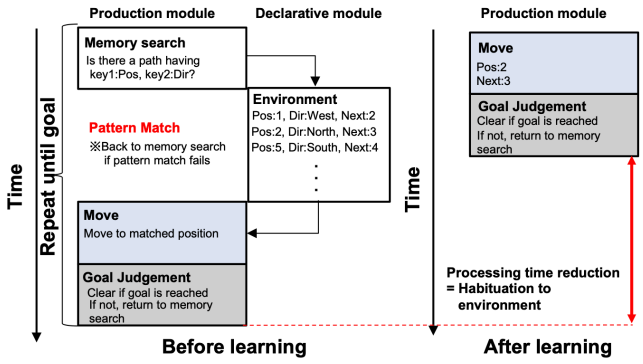

Figure 2: Example before and after learning using the *production compilation module*.

By applying this mechanism to the search of the maze environment, at first, the model often performs memory retrieval from the environmental map inside the declarative knowledge. As the task progresses, those memory retrievals becomes unnecessary. As a result, the model runs the tasks efficiently, and the frequency of pattern matching decreases due to the exhaustion of patterns in the environment.

Mechanism of the Task Continuation

Using the primitive functions presented so far, we propose our original mechanism of intellectual curiosity to determine whether to continue or to stop the task. Figure 3 illustrates a proposed mechanism of the continuation of a task in a general environment. At the start of each round, the model decides whether to continue or stop the task (conflict resolution between two rules). After it decides to continue the task, the model proceeds with the round by firing various rules (searching the map, etc.). When the model encounters a condition that ends the round, a new round is started, and the model decides whether to continue or stop the task again.

In the above process, the initial value of the utility of the *continue rule* is considered higher than that of the *stop rule*. At the beginning of the task, it can be assumed that humans intend to continue the task. The process of becoming bored from this initial state can be modeled by assigning a trigger of a negative reward to the rule that recognizes the end of each round. By triggering a negative reward at the end of the round, the utility of the *continue rule*, which have fired as a result of the previous conflict resolution, decreases, and the probability of firing the *stop rule* increases.

To prevent boredom and to consider the conditions for continuing environmental learning, a model of *intellectual curiosity*, namely *fun*, is required. If the model finds *fun* during the task, a positive reward is triggered, and the utility of

[1] Anderson (2007) made this claim at the introduction of the ACT-R function called *dynamic pattern matching*.

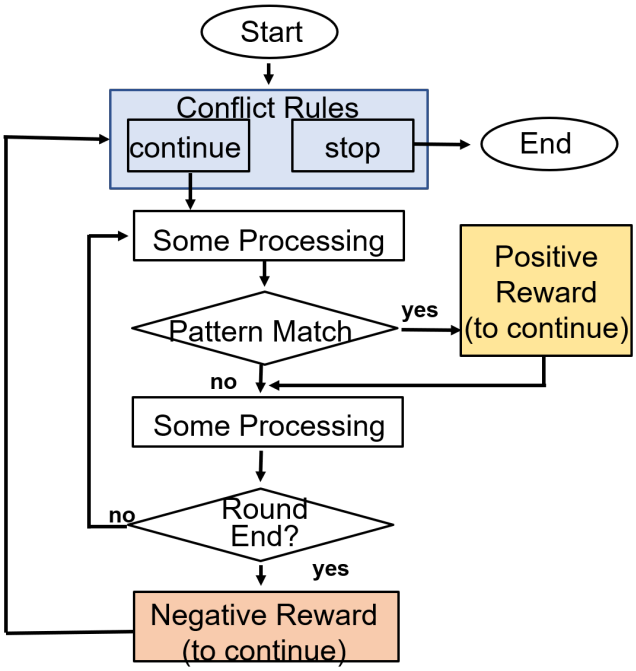

Figure 3: Flowchart of the task continuation model. Using "pattern matching" leads to a positive reward.

the *continue rule* maintains a high value. In this study, rules that trigger positive rewards are defined as rules that fire as a result of the successful retrieval of declarative knowledge in the task, such as remembering the map. The search for declarative knowledge requires pattern matching between the conditional clauses of the rule (the current situation) and the memory in declarative knowledge, and its success is consistent with Koster's definition of *fun* (finding patterns). However, this rule gradually becomes used to repeated execution; that is, the integration of rules occurs. After integration occurs, it becomes routine and cannot receive a reward. Then, the utility value of the *continue rule* decreases, and the *stop rule* fires. In short, long-term task continuation is achieved by keeping the model engaged in pattern matching between conditional clauses of production rules and declarative knowledge.

Implementation

The purpose of this study is to model intrinsic motivation by collecting primitive functions provided by ACT-R. For this purpose, it is necessary to implement the mechanism shown in the previous section on a specific task and observe its behavior. In this study, we select a path-planning maze task that has been used in many previous studies.

The Model of Maze Task

Our implementation of the ACT-R model for the maze search extends the memory-based strategy described in Reitter and Lebiere (2010) to include the mechanisms of task continuation (Figure 3).

Environment Figure 4 illustrates the maze in the present research. Reitter and Lebiere (2010) represented the maze environment as a topological map, consisting of a collection of cell IDs and links between cell IDs. In ACT-R, such a topological map is represented by a collection of chunks, and the model searches the environment, retrieving these chunks in the declarative memory.

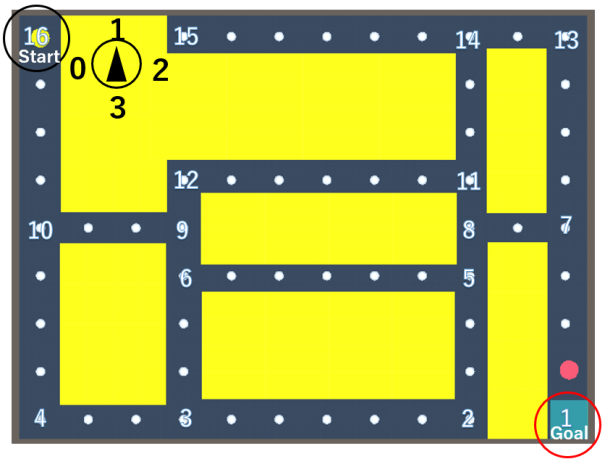

Figure 4: Maze environment

Searching Behaviours During the task, the model stores the current cell ID in the goal buffer. The model is initially located in #16 in Figure 4, and the model changes this state to #1 by retrieving a chunk stored in the declarative module. The chunk associated with the current position is requested, and the goal buffer is updated if the model retrieves such a chunk. This procedure is repeated until the model reaches the goal (#1).

Each time the goal is reached, the model labels all the chunks used in the current round as the correct path and stores these in the declarative module. From the next rounds, the model runs the task efficiently using these labeled chunks, following the method of the instance-based learning theory (Gonzalez et al., 2003).

If the model fails to retrieve the correct chunks, the model plans the path from the current position to the goal position using a heuristic search, namely a stochastic depth-first search (DFS). To realize backtracking used in a DFS, we implemented a stack structure using the imaginal module of ACT-R. Figure 5 depicts the stack function using chunks generated by this module. The push function in the stack is realized by generating a chunk that stores the name of the past chunk in the *ARG*1 slot. In addition, the pop function in the stack is realized by returning the *ARG*1 slot value to the past slot value. These generated chunks are stored in the declarative knowledge and can be retrieved later to realize the pop function. We implemented all these processes only through ACT-R production rules without defining any external functions written in other programming languages, such as LISP.

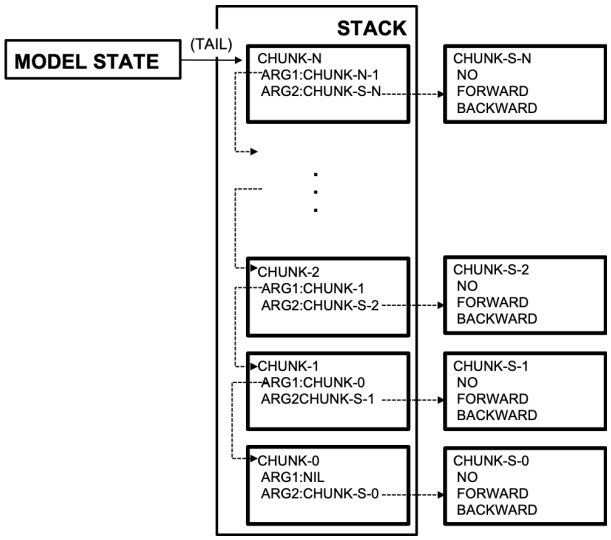

Figure 5: Stack structure of chunks implemented using the "imaginal module" of the ACT-R. During the search, the state of the model is dynamically stored (pushed) in the declarative module. The stored chunks are retrieved (popped) by following the dependencies noted in the slots of the chunks (dashed arrows).

Intrinsic Motivation in the Task In this simple maze task, we tried to observe how *fun* and *boredom* occur. In this model, *fun*, which is attached to pattern matching, is defined as remembering the correct path from the current situation to the goal. Specifically, the success of the DFS is defined as the attenuation of the motivation for continuing the task (positive rewards). In contrast, regardless of the success or failure of the goal search, a rule that fires at the end of the round is used as a trigger for a negative reward. The utility value of the *continue rule* decreases, as the negative trigger associated with the end of the round continues to occur without the rule expressing fun firing during the round. When the utility value of the *continue rule* falls below the utility value of the stop rule, the *stop rule* fires, and the task is terminated.

Simulation

Settings To confirm the behavior of our model of intrinsic motivation, we performed a simulation where the initial utility value of the *continue rule* was set to 10, and the initial utility value of the *stop rule* was set to 5. We also assigned the triggers of the negative reward ($r = 0$) to rules that recognized the end of the round (reaching goal #1 or recognizing that the time limit of each round has passed) and assigned the triggers of the positive reward to rules that included pattern matching. In this research, we select the *path finding rule* rule (DFS success) as the trigger of the positive reward, varying the value from 1 to 20 as the simulation conditions. For each condition of the positive reward value, the model runs the task 1000 times. In addition, we set the time limit of each round from 100 to 300 s. When the time limit was reached, the model resolved the conflict between the *continue* and *stop rules*.

The model also has rules that stochastically determine the directions to proceed (up, down, left, and right). The initial values of these utilities were also set to 10. Following Anderson et al. (2004), noise parameters were set as follows: ans (activation noise level) = 0.4 and, egs (production noise level) = 0.5.

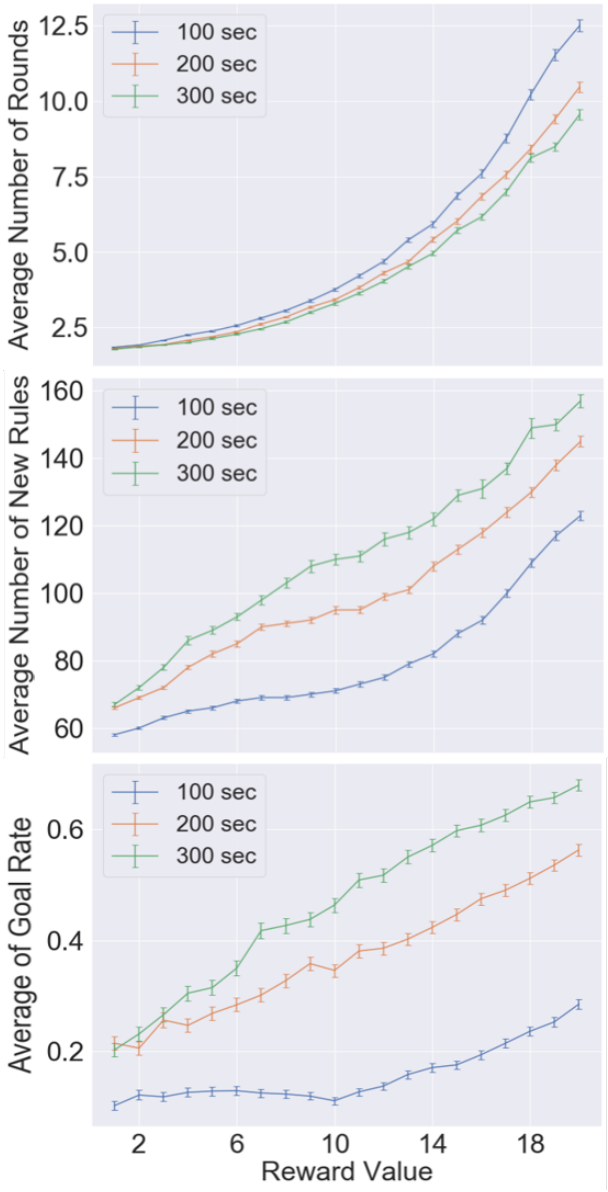

Figure 6: Results of the ACT-R model. Top: number of continued rounds; Middle: new rules generated by the production compilation; Bottom: goal rates of the model. The error bars in each graph represent the standard error.

Results Figure 6 displays the results of the simulation. From this figure, we observe that the reward generated by the pattern matching increased the number of continued rounds,

number of rules generated by the compilation module, and goal achievement rate. These results indicate that the implemented intrinsic motivation, which makes the model have a longer task continuation, leads to acquiring richer knowledge, resulting in better performance.

Comparison with Conventional Method

To further clarify the behavior of the proposed ACT-R model of intrinsic motivation, we compared it with a reinforcement learning agent that searched the maze environment in Figure 4. In this simulation, we used the algorithm of IMRL (Singh et al., 2005). At each time point, the reinforcement learning model is located on one of the numbered positions in the map, and it moves to up, left, down, or right. The selection of the direction is controlled with IMRL with ε-greedy ($\varepsilon = 0.2$, $\gamma = 0.9$, $\alpha = 0.2$):

$$Q(s,a) \leftarrow Q(s,a) + \alpha[\acute{r_i} + \acute{r_e} + \gamma max Q(\acute{s},\acute{a}) - Q(s,a)] \quad (1)$$

Equation 1 indicates the updating of the Q value of the present model, where r_e represents a reward from the external environment, and r_i represents a reward from the internal environment. The model receives $r_e = -1$ when it fails to remember the path, whereas it receives $r_e = 0$ when it successfully remembers the path. The model also obtains $r_e = 10$ when it reaches the goal. Contrary to r_e, r_i is determined with Equation 2.

$$r_i = -\tau(1-p)\log(1-p) \quad (2)$$

where p represents the transition probability with the Q value. From this equation, the internal reward r_i is determined as the entropy of the probability of the complementary event with respect to p. In addition, τ is the coefficient for the reward value used in the simulation. A larger τ indicates a greater intrinsic motivation. In this study, we manipulated this value from 0.34 to 0.74. For each condition of the τ value, the model runs the task 1000 times. We manipulated the number of steps of movement in each round (100, 125, and 150 steps). When the step limit was reached, the model chose to quit the task or to continue the task by comparing the summation of the obtained internal reward (r_i) with the given threshold ($th = 5$).

Figure 7 indicates the result of the simulation of the reinforcement learning model. Similar to the ACT-R model, the internal reward increases the number of rounds. However, in contrast to the ACT-R model, it slightly decreases the goal rate. That is, the reinforcement learning model with high intrinsic motivation does not learn to achieve the goal but learns to explore the environment. This behavior might be changed by modulating the balance between r_i and r_e in Equation 1 or by designing the maze environment carefully to stimulate curiosity, as suggested by Burda et al. (2018). However, our ACT-R model could learn the environment without such careful parameter modulations or environmental design. Therefore, from this simulation, we can claim the advantage of our model in representing intrinsic motivation.

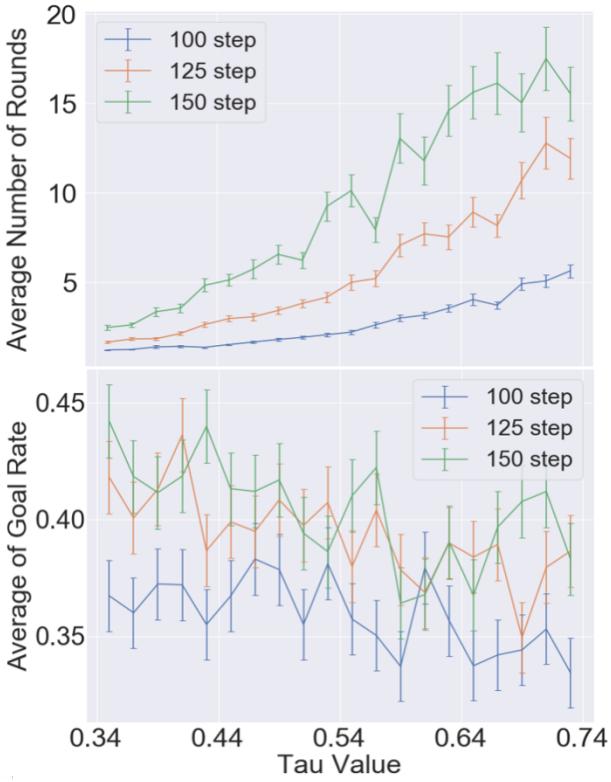

Figure 7: Results of the reinforcement learning model. Top: number of continued rounds; Bottom: goal rates. The error in each graph is the standard error.

Conclusion

The purpose of this study was to construct a model of intrinsic motivation by accumulating primitive cognitive processes provided by ACT-R. To achieve this goal, we assumed that the mechanism of pattern matching represents the source of intellectual curiosity, namely fun. Thus, with the success of pattern matching, the model maintains high intrinsic motivation for task continuation. In contrast, by skipping the pattern matching with compilation mechanisms, the model 'tires' of the task and eventually stops.

From the simulation results presented in Figures 6 and 7, we consider that our model has an advantage in learning new environments. The model uses both the utility module and instance-based learning (memorizing the correct path to the goal; Gonzalez et al., 2007). Such a combination of several learning algorithms might help balance the intrinsic and extrinsic rewards in the current maze task.

However, the result in the previous section does not indicate that the conventional reinforcement learning cannot achieve the same learning as ACT-R. The model presented by Singh et al. (2005) included the mechanism called option, which summarizes low-level actions into abstract-level units (Sutton, Precup, & Singh, 1999) and indicated the process of moving up using abstract *option* as the model learned the environment. This mechanism has a commonality with

the compilation module in ACT-R. Schmidhuber (2010) also pointed out that such a compression mechanism is the same as the prediction mechanism, which leads to the emotional process of fun and boredom. We need to further explore the relationship between such models of reinforcement learning and the proposed model.

In addition to the efficacy of environmental learning, the expression of the internal reward of our model has an advantage compared to previous studies. In our model, we did not explicitly divide the internal and external rewards in the equation, but the effect of intrinsic motivation is represented in the existing mechanism of ACT-R. We consider that this approach has an advantage because it is based on the theory of human cognition, it is related to the existing learning research, and it saves unnecessary factors in the theory.

In future studies, we need to compare the model of intrinsic motivation with human data. As a model of human cognition, behavior presented by conventional reinforcement learning might not be wrong. During search tasks, people often forget the goals and decrease in performance. Such irrational behavior might also relate to *computational psychiatry* (Huys, Maia, & Frank, 2016).

We also need to model the optimal level of motivation (Yerkes & Dodson, 1908). In this study, the model statistically determines the initial utility value of the *continue rule* to focus on the decay process of intellectual curiosity. The process up to the optimal level, obtaining intrinsic motivation for the target environmental learning, is not modeled. Therefore, by constructing a model representing such a process, we can explore more detailed conditions of task continuation, especially those before the model reaches optimal levels.

Acknowledgment

This research was supported by JSPS KAKENHI Grant Numbers 17H05859, 20H05560, 20H04996.

References

Anderson, J. R. (2007). *How can the human mind occur in the physical universe*. Oxford Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036.

Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., & Efros, A. A. (2018). Large-scale study of curiosity-driven learning. *arXiv preprint arXiv:1808.04355*.

Dancy, C. L., Ritter, F. E., Berry, K. A., & Klein, L. C. (2015). Using a cognitive architecture with a physiological substrate to represent effects of a psychological stressor on cognition. *Computational and Mathematical Organization Theory*, *21*(1), 90–114.

Fu, W.-T., & Anderson, J. R. (2006, 6). From recurrent choice to skill learning: A reinforcement-learning model. *Journal of Experimental Psychology. General*, *135*, 184-206.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635.

Huys, Q. J., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, *19*(3), 404.

Juvina, I., Larue, O., & Hough, A. (2018). Modeling valuation and core affect in a cognitive architecture: The impact of valence and arousal on memory and decision-making. *Cognitive Systems Research*, *48*, 4 - 24.

Koster, R. (2004). *Theory of fun for game design*. Paraglyph Pr.

Kotseruba, I., & Tsotsos, J. K. (2020). A review of 40 years of cognitive architecture research: Focus on perception, attention, learning and applications. *AI Review*, *53*, 17-94.

Manoury, A., Sao, M. N., & Cédric, B. (2019). Hierarchical affordance discovery using intrinsic motivation. In *In Proceedings of the 7th International Conference on Human-Agent Interaction (HAI 19)* (pp. 186–193).

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... others (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.

Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 16–17).

Reitter, D., & Lebiere, C. (2010). A cognitive model of spatial path-planning. *Computational and Mathematical Organization Theory*, *16*(3), 220–245.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, *2*(3), 230–247.

Singh, S., Barto, A. G., & Chentanez, N. (2005). Intrinsically motivated reinforcement learning. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems 17* (pp. 1281–1288). MIT Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, *112*, 181–211.

Taatgen, N. A., & Lee, F. J. (2003). Production compilation: A simple mechanism to model complex skill acquisition. *Human Factors*, *45*(1), 61–76.

van Vugt, M. K., & van der Velde, M. (2018). How does rumination impact cognition? A first mechanistic model. *Topics in Cognitive Science*, *10*(1), 175–191.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. King's College, Cambridge.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482.