Paper:

A Basic Research to Develop a Method to Classify Game Logs and Analyze them by Clusters

Akinobu Sakata[*1], Takamasa Kikuchi[*4], Ryuichi Okumura[*5], Masaaki Kunigami[*2], Atsushi Yoshikawa[*3], Masayuki Yamamura[*1], and Takao Terano[*6]

[*1]Tokyo Institute of Techonology, Yokohama, Japan
E-mail: {sakata.a.aa, yoshikawa.a.aa, yamamura.m.aa}@m.titech.ac.jp
[*2]Tokyo Institute of Techonology, Yokohama, Japan
E-mail: mkunigami@gakushikai.jp
[*3]Tokyo Institute of Techonology, Yokohama, Japan
E-mail: at_sushi_bar@dis.titech.ac.jp
[*4]Keio University, Yokohama, Japan
E-mail: takamasa_kikuchi@keio.jp
[*5]Mitsubishi Research Institute,Inc., Tokyo, Japan
E-mail: okuty@mri.co.jp
[*6]Chiba University of Commerce, Ichikawa, Japan
E-mail: terano@cuc.ac.jp
[Received 00/00/00; accepted 00/00/00]

Abstract. The objective of this study is to show that it is possible to classify a set of simulation logs of a gaming simulation into several clusters, and to reveal the characteristics of the gaming results by analyzing them in clusters. This research is a stepping stone toward the development of a new analysis method that is somewhere between the approaches used in previous research, which statistically analyze all simulation logs of a gaming simulation and the analysis approach that closely observes a specific simulation log. In the former approach, when the simulation logs are divided into several clusters by characteristics, the values of statistical indicators extracted from all simulation logs are not enough to fully grasp the characteristics of the gaming results. The latter approach requires a lot of time and effort to analyze, and can only tell us about a small portion of the simulation logs. The analysis method we aim to develop compensates for the shortcomings of both approaches. We collected a large number of logs in an experiment using a gaming simulation we developed, classified the log set, and checked the distribution of log features in each cluster. As a result, we showed that the distribution of feature values among clusters is different.

Keywords: Gaming and Simulation, Agent-Based Simulation, Log Cluster Analysis

1. Introduction

The objective of this study is to show that the characteristics of gaming outcomes can be clarified by classifing the play logs of gaming simulations and then analyzing the play logs in groups. A gaming simulation is a simulation in which a human player participates in the simulation situation as a player and is controlled by the decisions of the player[2].

The first is the general scientific approach of designing an experiment based on statistics, collecting data, and analyzing the data to clarify the relationship between the explanatory variables and the explained variables. In this study, this analytical approach is referred to as the macroscopic analytical approach. The second analytical approach is to focus on a specific play log to track players' decisions and actions during the game and observe them in detail (e.g. [8]). In this study, we refer to this analytical approach as the micro-analytical approach.

The macroscopic and microscopic analysis approaches described above have been used in research areas outside the field of gaming simulation. Matsuda et al. proposed to improve the efficiency of experiments using agent-based simulation by utilizing the design of experiments method[7]. Kobayashi et al. proposed a method to analyze virtual business cases acquired by agent-based simulation by comparing them with business cases written based on real businesses[5].

On the other hand, research using agent-based models has proposed an analysis method that is neither the macro nor micro analysis methods described above, but is positioned in between the two[9]. Tanaka et al. used hierarchical clustering to classify the logs (records of agents' actions, decisions, and states in chronological order) acquired through organizational simulation into multiple clusters, and identified the factors of classification through decision tree analysis. There are a few studies that have developed this method (e.g. [3][4]).

Inspired by the analysis method proposed by Tanaka and colleagues, we have been investigating the possibility of applying it to the analysis of gaming simulation play logs. To the best of our knowledge, there has been no case of classifying a set of simulation logs of a gaming simulation and analyzing the simulation logs by cluster. Therefore, we attempted to classify the simulation logs of gaming simulations using a hierarchical clustering method, and compared the distribution of features among the clusters to see if they had different characteristics from each other. If this attempt is successful, it will bring us one step closer to developing a new method of analyzing the simulation logs of gaming simulations.

In this study, we used a modified version of the Shin-Life Career Game[8], a gaming simulation for career education, for our experiments. In this game, players experience a hypothetical worker's life. The player is equipped with four resources (money, ability, time, and health) and repeatedly distributes the resources among five activities (regular employment, freelance work, simple work, study, and leisure) as in a life game[1].

In this experiment, we first made a software agent, a virtual player who randomly selects one of 11 actions in each turn, play the game 1,000 times in each condition under a total of 120 conditions. The simulation logs were then clustered by condition using a hierarchical clustering method, with the money resource variable and the ability resource variable at the end of the game as features. Finally, we compared the distribution of the data in each cluster under each condition. After the experiment, the simulation logs were clustered by condition using the hierarchical clustering method.

2. Related Work

In this section, we describe the log cluster analysis method developed by Tanaka et al. Tanaka et al. have developed a method for categorizing all logs of a simulation into a set of logs according to the similarity of the structure of the simulation model, and then comparing and analyzing the log clusters with each other.

The analysis method proposed by Tanaka et al. consists of two components: the classification of simulation logs and the identification of data classification factors using decision tree analysis. First, clustering of the simulation logs is performed to obtain multiple clusters. The properties of each cluster are analyzed, and the clusters are compared with each other to better understand the simulation results. Next, the simulation logs are subjected to decision tree analysis to extract knowledge that reveals which elements of the simulation logs are divided into clusters and under what conditions.

3. Methodology

3.1. The Shin-Life Career Game

In this section, we describe a modified version of the Shin-Life Career Game.

3.1.1. Resource variables

In this section, we describe the resources that players have in the modified version of the Shin-Life Career Game. Players have four kinds of resources (money resources, ability resources, time resources, and health resources) and are allowed to freely allocate each resource to five kinds of activities (Permanent work (PW), Freelance work (FW), Simple work (SW), Learning (LN), and Leisure (LS)) according to their intentions. When resources are supplied to each activity, a reward (some kind of resource) is given to the player according to the amount of resources allocated. The relationship between the resources input to each activity and the resources output from the activity is described by the MATH model[6] (see figure 1). In the following, we describe the resource variables.

The variable of money resource corresponds to an asset in the real life. Human beings use monetary assets to support their lives. The variable of money resource has extensive property and does not have an upper limit.

The variable of ability resource corresponds to knowledge and skills for work in the real life. Human beings use their skills and knowledge to engage in labor and get paid for it. The variable of ability resource has intensive property and does not have an upper limit.

The variable of time resource corresponds to time for real life. Human beings live their lives by spending their time in a variety of activities. The variable of time resource has extensive property and an upper limit.

The variable of health resource corresponds to health or physical fitness for real life. Health and physical fitness are the foundation of real life. The variable of health resource has intensive property and an upper limit.

3.1.2. Activity

In this section, we describe the activities of agents. Players in the modified version of the Shin-Life Career Game can allocate resources to five types of activities: permanent work (PW), freelance work (FW), simple work (SW), learning (LN), and leisure (LS). In the following, the definitions and characteristics of the above five types of activities are described. Note that activities whose contents overlap with the MATH model of the original gaming simulation (the features of PW, FW, SW, and LN) are not described in detail (see [8]).

PW is a work style in which workers are employed by an organization until they reach retirement age and receive remuneration for their labor. The characteristics of PW are described below. First, the remuneration for PW is stable. Second, the remuneration for PW is higher than

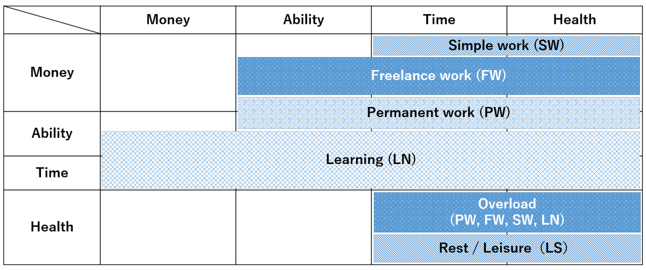


Fig. 1. : This figure shows the relationship between the resources input to each activity and the resources output (A modified version of the Shin-Life Career Game).

that for SW. Third, if a player's ability resources increase, the reward for PW will be higher. Fourth, engaging in PW increases the ability resources of players. Fifth, a player has to provide a certain amount of time resources for PW. Sixth, a player is not able to decide the amount of time resources to be allocated to PW at will.

FW is a way of working that is independent of a particular organization and is paid by providing expertise and skills to a contracted party. FW has the following characteristics. First, the income of workers who engage in FW is unstable. Second, as a player's ability resources increase, the reward for FW increases. Third, a player's ability resources do not increase when the player engages in FW. Fourth, a player is free to decide the amount of time resources to be allocated to FW.

SW is a form of work in which workers provide their time to their employers or contractors and are paid for it. In this game, SW is the kind of work that manual workers do in the real world, such as part-time jobs, day labor, and gig work, which do not require any special skills or qualifications. SW has the following characteristics. First, the money resource, which is the reward for labor, increases in proportion to the amount of time resource the player allocates to SW. Second, the amount of a player's ability resource do not affect the amount of compensation for SW. Third, the income of a player who engages in SW is unstable. Fourth, a player's ability resource does not increase when the player engages in SW. Finally, a player has the flexibility to adjust his/her working hours to engage in SW.

LN is the act of taking extra time to develop competencies in order to nurture one's work capacity. Learning in this game means that working people acquire new knowledge and skills at universities and attend courses and workshops to receive specialized knowledge.

In this study, we assume that when agents choose to engage in these activities (PW, FW, SW, and LN), health resources are lost in return. We refer to this phenomenon as overload (OL). In general, overwork in labor and study impairs human health, which in turn leads to lower labor productivity and reduced effectiveness in learning.

Finally, we will discuss a novel activity, leisure (LS). LS is an activity that workers set aside in order to recover their physical and mental health. In general, workers use their leisure time for rest and recuperation to maintain their health condition.

Based on each of the above characteristics, we examined a model that describes the relationship between the resource variables input to each activity and the resource variables output from each activity. The amount of remuneration for PW is considered to increase monotonically in proportion to the product of the amount of wage per unit of time, the length of working hours, the level of worker's ability and the degree of influence of the health condition (see equation (1)). The ability of PW is expected to increase monotonically in proportion to the product of the learning effect per unit of time, the length of working hours, the level of ability, and the degree of influence of the health condition (see equation (2)). The amount of remuneration for FW increases monotonically in proportion to the product of the length of working hours, the degree of worker's ability and the degree of influence of the health condition, but there is uncertainty in the income side (see equation (3)). The amount of compensation for SW is expected to increase monotonically with the product of the length of working hours and the degree of influence of the health condition (see equation (4)). The degree of growth of the worker's ability is expected to increase monotonically in proportion to the product of the learning effect per unit of time, the length of working hours, the cost of learning, the level of working ability, and the degree of influence of the health condition (see equation (5)). The degree of deterioration in health status is expected to increase monotonically in proportion to the product of the degree of deterioration per unit time and the activity time (see equation (6)). The degree of recovery of the health state is considered to increase monotonically in proportion to the product of the degree of recovery and the leisure time per unit time (see equation (7)). The degree of influence of the health condition is the effect of the past health condition after a time delay (see equation (8)). Based on the above, equations (1)-(8) were constructed. The values of each resource variable are also updated in equations (9)-(12). Refer to Table 2 for the variables and constants that make up each equation.

$$I_{PW}(t) = c_{PW} \times A(t-1) \times T_{PW}(t) \quad (1)$$

$$G_{PW}(t) = \gamma_H(t) \times c_{EF_PW} \times T_{PW}(t) \times A(t-1) \quad (2)$$

$$I_{FW}(t) = \gamma_H(t) \times \varepsilon_{FW} \times c_{FW} \times T_{FW}(t) \times A(t-1) \quad (3)$$

$$I_{SW}(t) = \gamma_H(t) \times c_{SW} \times T_{SW}(t) \quad (4)$$

$$G_{LN}(t) = \gamma_H(t) \times c_{EF\_LN} \times T_{LN}(t) \times \sqrt{M_{LN}(t)} \times A(t-1) \quad (5)$$

$$H_{BRDW}(t) = c_{BRDW} \times (T_{PW}(t) + T_{FW}(t) + T_{SW}(t) + T_{LN}(t)) \quad (6)$$

$$H_{RCV}(t) = c_{RCV} \times T_{LS}(t) \quad (7)$$

$$\gamma_H(t) = \begin{cases} 1 & (1 \le t \le 2) \\ \frac{H(t-2)}{H_{MAX}} & (2 < t \le 40) \end{cases} \quad (8)$$

$$M(t) = M(t-1) + I_{PW} + I_{FW} + I_{SW} - M_{LN} \quad (9)$$

$$A(t) = A(t-1) + G_{PW} + G_{LN} \quad (10)$$

$$T(t) = T_{MAX} - (T_{PW} + T_{FW} + T_{SW} + T_{LN} + T_{LS}) \quad (11)$$

$$H(t) = H(t-1) + H_{RCV}(t) - H_{BRDW}(t) \quad (12)$$

3.1.3. Happening

A modified version of the Shin-Life Career Game has a function which happens an economic depression. The economic depression forces players to change their decisions for resources allocation. The amount of time resources allocated to PW, FW, and SW is forcibly reset to the minimum amount, and all surplus time resources are allocated to LS. In this study, the minimum amount of time resources that a player can allocate to PW is set to 70, and the minimum amount of time resources that a player can allocate to FW and SW is set to 0. Permanent workers

Table 1. : Variables and constants in equations

Sign	Variable/Const	Description
$M(t)$	variable	Player's money resource as of round t ($0 \leq M(t)$)
$A(t)$	variable	Player's ability resource as of round t ($0 \leq A(t)$)
$T(t)$	variable	Player's time resource as of round t ($0 \leq T(t) \leq T_{MAX}$)
$H(t)$	variable	Player's health resource as of round t ($0 \leq H(t) \leq H_{MAX}$)
$I_{PW}(t)$	variable	Reward for PW in round t (money resource)
$I_{FW}(t)$	variable	Reward for FW in round t (money resource)
$I_{SW}(t)$	variable	Reward for SW in round t (money resource)
$G_{PW}(t)$	variable	Reward for PW in round t (ability resource)
$G_{LN}(t)$	variable	Reward for LN in round t (ability resource)
$H_{RCV}(t)$	variable	Reward for LS in round t (health resource)
$H_{BRDW}(t)$	variable	Penalty for activities (except LS) in round t (health resource)
$M_{LN}(t)$	variable	Money allocated to LN in round t by a player (money resource)
$T_{PW}(t)$	variable	Time spent working as a permanent worker in round t (time resource)
$T_{FW}(t)$	variable	Time spent working as a freelance worker in round t (time resource)
$T_{SW}(t)$	variable	Time spent working as a simple worker in round t (time resource)
$T_{LN}(t)$	variable	Time spent developing ability in round t (time resource)
$T_{LS}(t)$	variable	Time spent recovering health in round t (time resource)
$\varepsilon_{FW}(t)$	variable	A random number generated according to a continuous distribution whose probability density function is constant on a finite interval $[\alpha,\beta]$ and zero outside the interval.
$\gamma_H(t)$	variable	Influence of health status on performance of each activity ($0 \leq \gamma_H(t) \leq 1$)
T_{MAX}	constant	Initial value of a player's time resource (time resource)
H_{MAX}	constant	Initial value of a player's health resource (health resource)
c_{PW}	constant	Reward per unit time for PW
c_{FW}	constant	Reward per unit time for FW
c_{SW}	constant	Reward per unit time for SW
c_{LN}	constant	Reward per unit time for LN
c_{BRDW}	constant	Amount of health resources lost per unit of time
c_{RCV}	constant	Amount of health resources recovered per unit time

are less likely to be dismissed because of the economic downturn, and thus receive compensation corresponding to the least amount of time resources allocated. Therefore, if the agent chooses PW, he can secure some income even if an economic recession occurs. However, agents who choose FW and SW will have their allocated amount of time resources returned to zero and experience a significant temporary decrease in income.

3.2. Agent

In this study, we employ software agents as players in a gaming simulation. In order to simplify the experiment, we created a software agent that makes decisions according to the rule of randomly selecting one resource allocation action from a total of eleven resource allocation actions in each round.

3.3. Instructions for Gameplay

As a player play the game, he/she has to make decisions about resource allocation in each round. At the beginning of each round, the player gets the opportunity to check the information about the game. Here, the player

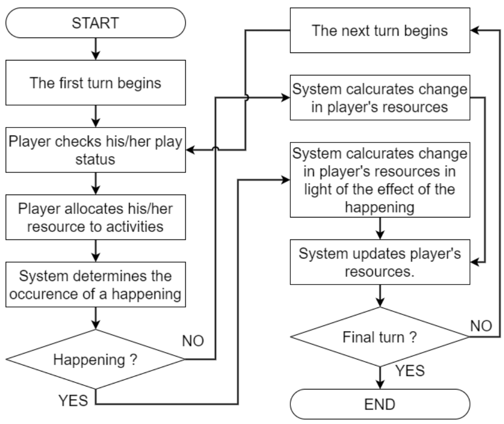


Fig. 2. : Flowchart for gameplay

is allowed to view the history of the amount of resources he/she has, the history of his/her past actions and results, and the history of economic recession event. Based on the information obtained in this stage, the player considers the policy of resource allocation. Next, the player allocates resources to each activity. After the resource allocation is completed, the game system updates the player's resource variables. At this time, if the game system considers that

Table 2. : Resource allocation actions that can be selected by software agents

Action	Objective	Allocated Resources (%)†					Outcome‡		
		T_{PW}	T_{FW}	T_{SW}	T_{LN}	T_{LS}	M	A	H
A	Money-making	RD_{PW}		100-T_{PW}			+	+	-
B			100				+		-
C				100			+		-
D	Ability Development	RD_{PW}			100-T_{PW}		+	+	-
E			50		50		+/-	+	-
F					RD_{LN}	100-T_{LN}	-	+	-
G					100		-	+	-
H	Enjoying Leisure	RD_{PW}				100-T_{PW}	+	+	+/-
I			50			50	+		+/-
J					100-T_{LS}	RD_{LS}	-	+	+/-
K						100			+/-

† RD_{PW} is randomly determined in increments of 10 from 70 to 100. RD_{LN} and RD_{LS} are also randomly determined in increments of 10 from 50 to 100.

‡ ”M” stands for money resource, ”A” for ability resource, and ”H” for health resource. ”+” means resource increase, ”-” means resource decrease, and ”+/-” means one of the two can happen.

an economic depression event has occurred, the effect is reflected in the calculation results. Finally, the results of the calculation are fed back to the player, and the round ends. Then the game moves to the next round. The game rounds are repeated a predetermined number of times, and then the game ends.

3.4. System

The software for simulation is written in python. The simulation was carried out using Intel(R) core(TM) 4600U CPU @ 2.10GHz PC with 16GB RAM, Windows10 Pro, 64bit OS.

3.5. Procedures

The software agent is given 40 opportunities to make a decision per game (forty rounds). Each game was conducted by selecting one condition (120 conditions in all) from among all combinations of values for the incidence of economic recession (0, 10, 20, 30, 40, 50 percent), the amount of health capital lost per unit time (c_{BRDW}: 0.05, 0.1, 0.15, 0.2), and the amount of health capital recovered per unit time (c_{RCV}: 0.1, 0.2, 0.3, 0.4, 0.5) (120 conditions in total). We will check the effect of these factors on the amount of money resources and ability resources a player has at the end of the game. The game was repeated 1000 times for each condition.

4. Results and Discussion

First, the simulation logs were organized by experimental condition. Next, the values of the money and ability resource variables at the end of the game were used as features, and the Euclidean distance was used to calculate The Euclidean distance was used to calculate the distance between simulation logs with the same experimental conditions. After that, The simulation logs were then

Table 3. : Values of the parameters of the MATH model in this experiment

Parameter	Value	Parameter	Value
M_0	0	c_{SW}	5.5
A_0	1	c_{LN}	2.0×10^{-4}
T_{MAX}	100	α	2.5×10^{-1}
H_{MAX}	100	β	2.0
c_{PW}	5.5		
c_{FW}	5.5		

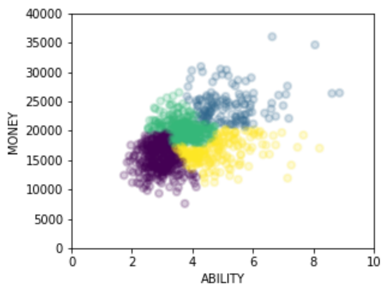
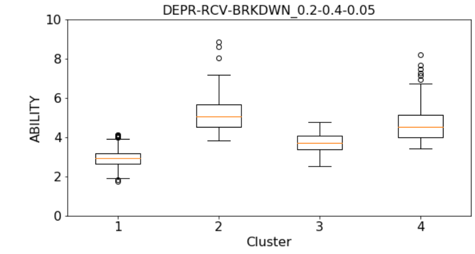
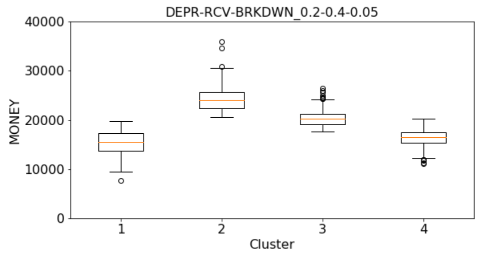


Fig. 3. : This figure shows the data from each play log plotted on a two-dimensional plane, with the money and ability resource variables at the end of 40 rounds on the y-axis and x-axis, respectively. The figure shows that the total of 1000 samples was classified into four clusters.

classified using a hierarchical clustering method based on the Ward method. The number of clusters was determined according to the value of the silhouette coefficient. Since there were 120 experimental conditions in total, only one condition was selected from among them (incidence of economic recession: 20 percent, amount of health resources lost per unit time: 0.05, amount of health resources recovered per unit time: 0.4). The results of the analysis (scatter plots (See figure 3), box-and-whisker



(a) The distribution of the values of the ability resource variable by cluster.



(b) The distribution of the values of the money resource variable by cluster.

Fig. 4. : This figure shows the results of the comparison of the distribution of the values of the money resource variable and the ability resource variable at the end of 40 rounds in a box-and-whisker diagram. However, an interesting discrepancy was observed between Class 3 and Class 4, where the value of the money resource variable was larger in Class 3 than in Class 4, even though there was no difference in the value of the ability resource variable.

plots (See figure 4a, figure4b) showing the distribution of the values of the money resource variable and the ability resource variable) were published in this journal.

The gaming simulation used in this study is designed in such a way that the more ability resources a player has, the more money resources he can earn through regular employment or freelance labor. Therefore, a software agent with a higher value of the ability resource variable in the final round should also have a higher value of the money resource variable. However, this is not necessarily the case in the experimental results. First, Figure 4b shows that the value of the ability resource variable in cluster 3 is lower than that in cluster 4. On the other hand, in Figure 4a, the value of the money resource variable in cluster 3 is lower than the value of the money resource variable in cluster 4. This is inconsistent with the properties of the game model described above.

We believe that the interesting patterns described above emerged because the software agents discovered some special paths in the process of playing the game. For example, the software agents with simulation logs belonging to cluster 4 spent less resources on learning than the software agents belonging to cluster 3, but they may have been lucky enough to have multiple opportunities to earn high rewards for their freelance work. The third analytical approach that we aim to develop may be useful in identifying the conditions under which gaming paths diverge in such gaming simulations.

5. Conclusion

The objective of this study is to show that the characteristics of gaming results can be clarified by analyzing the play logs of gaming simulations on a cluster basis after typifying the simulation logs. To achieve this objective, we made software agents play the Shin-Life Career Game gaming simulation many times, classified the simulation logs, and clarified the distribution of money resource variable and ability resource variable in each cluster. As a result, we found a large number of simulation logs with characteristics that contradict the results expected from the characteristics of the model, such as differences in the values of the money resource variable even when the values of the ability resource variable at the end of the game were equal. These results indicate that the simulation logs of gaming simulations can be classified into multiple clusters, and that each cluster has different characteristics. Based on this, we intend to explore the effectiveness of the approach of classifying the data from gaming simulations into clusters for analysis in the future.

References:

[1] J. Ferrara. Games for Persuasion: Argumentation, Procedurality, and the Lie of Gamification. *Games and Culture*, Vol. 8, No. 4, pp. 289–304, jul 2013.

[2] C. S. Greenblat. *Designing games and simulations: An illustrated handbook*. SAGE Publications Inc., 1988.

[3] Takamasa Kikuchi, Masaaki Kunigami, Hiroshi Takahashi, Masahiro Toriyama, and Takao Terano. Explaining Log Data of Agent-Simulation Results with Managerial Decision-Making Description Model. *Simulation & Gaming*, Vol. 29, No. 1, pp. 36–48, 2019.

[4] Takamasa Kikuchi, Masaaki Kunigami, Hiroshi Takahashi, Masahiro Toriyama, and Takao Terano. Classification of Simulation Results for Formal Description of Business Case. *Journal of the Japan Society for Management Information*, Vol. 29, No. 3, pp. 199–214, 2020.

[5] Tomomi Kobayashi, Satoshi Takahashi, Masaaki Kunigami, Atushi Yoshikawa, and Takao Terano. Analyzing Organisational Innovation and Deviation Phenomena by Agent Based Simulation and Case Design. In *Proceedings of the 9th International Conference on Innovation & Management*, pp. 778–791, 2012.

[6] Masaaki Kunigami, Ryuichi Okumura, Takamasa Kikuchi, Akinobu Sakata, and Takao Terano. Agents' Resources Allocation and Growth Modeling for Social and Organizational Agent Simulations. In *JSAI Special Interest Group on Business Informatics (SIG-BI #17)*, pp. 55–58, Online, 2021.

[7] Hiroyasu Matsushima, Takeshi Uchitane, Junpei Tsuji, Tomohisa Yamashita, Nobuyasu Ito, and Itsuki Noda. 実験計画法による実験数削減と有意なパラメータ探索の避難シミュレーション分析への適用 (Applying Design of Experiment based Significant Parameter Search and Reduc-ing Number of Experiment to Analysis of Evacuation Simulation). *Transactions of the Japanese Society for Artificial Intelligence : AI*, Vol. 31, No. 6, pp. AG–E_1, 2016.

[8] Akinobu Sakata, Takamasa Kikuchi, Ryuichi Okumura, Masaaki Kunigami, Atsushi Yoshikawa, Masayuki Yamamura, and Takao Terano. The Shin-Life Career Game: Pursuing Your New Life Style through Gaming Simulation. In Takao Terano and Susan Gauch, editors, *The Thirteenth International Conference on Information, Process, and Knowledge Management (eKNOW 2021)*, pp. 14–20, Nice, France, 2021.

[9] Yuji Tanaka, Masaaki Kunigami, and Takao Terano. What can be Learned from the Systematic Analysis of the Log Cluster of Agent Simulation. *Simulation & Gaming*, Vol. 27, No. 1, pp. 31–41, 2018.