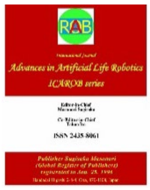
SUGISAKA MASANORI

Journal of Advances in Artificial Life Robotics
Vol. 1(3); December (2020), pp. 138–143
ON LINE ISSN 2435-8061; ISSN-L 2435-8061
https://alife-robotics.org/jallr.html



Research Article

Respawning point recommendation by TD-Learning as a content generation of FPS video game-like E-learning

Masao Kubo, Takeshi Ueno, Hiroshi Sato
Department of Computer Science, National Defense Academy, Hashirimizu-10-20, Yokosuka, Kanagawa 239-8686, Japan

ARTICLE INFO

Article History

Received 26 November 2019
Accepted 23 August 2020

Keywords

Content generation
Reinforcement learning
E-Learning
Game AI
Video game
Virtual reality

ABSTRACT

In this paper, we propose an approach to customize the E-learning of video game-like by trial and error. A virtual training environment is getting to be common in military training, however, it is still underway to use it as a self-learning tool because of a lack of suitable training curricula for each trainee. First Person Shooting game (FPS) environment which is adequate for such the training, but a lot of characters and objects there which can be considered as the customizing point may cause combinatorial problems in traditional approaches. We show our method based on respawning point can present tasks to trainees by reinforcement learning and they can reach the goal faster than other content generation methods.

1. Introduction

The E-learning is getting important year by year and so a better quality E-learning system is required. One of the main issues of E-learning is the quality of recommendation of tasks and contents to be learned which fits the student's skill. When the computer-selected tasks are unfitted to the students it is difficult to be a tool for self-learning because extra supports by human are required.

E-learning has already been incorporated into the curriculum in the US military [1]. In Japan the virtual training environment is getting to be common for example, training of gun shooting, vehicle control, etc.

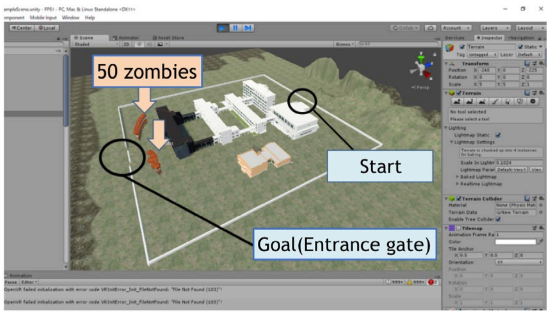


Fig. 1. Field of the FPS like E-Learning

and the corresponding human skilled trainers customize trainee's learning content. Especially FPS video game-like virtual training is adequate for training for decision making in real-time also it has potentially a wider degree of freedom to customize the content for trainees than traditional E-learning. Generally speaking, a human trainer can customize the tasks well, which trainee can acquire its wishing skill efficiently because the trainer

Corresponding author's E-mail: masaok@nda.ac.jp, hsato@nda.ac.jp URL: https://www.mod.go.jp/nda/

knows the domain well. However, new domains to be handled are increasing year by year, so a lack of such expert trainers may become a serious problem.

So, in this paper automatic efficient learning task generation of FPS video games like virtual learning environment is discussed. As we described above FPS games are adequate for training for decision making in real-time, meanwhile few researches have been executed as E-learning media so far.

An FPS game has consisted of a series of scenes which include many characters, creatures, objects, etc. [2]. All of them can be considered as the control points of difficulty of the E-learning, however, this approach is based on the deployment of these characters may cause combinatorial problems. In this paper, we propose another control point which does not depend on such game scene design.

There are some arcade games that a player has to start its game from the beginning whenever it fails the game along the way. It may be a good trait as entertainment, meanwhile, it is inefficient as a learning experience of the point of view because skillful trainees still have to play scenes of the game which are too easy to obtain knowledge. Therefore, in this paper, we aim to adjust starting scenes after the mistakes of trainees to shorten a time to solve a game. Usually, the scenes are called a respawning point in [3] and we use this term in the following pages of paper. We propose respawning points as the difficulty control of the E-learning. The proposed method recommends respawning points to students which can be expected to resolve the task as soon as possible. TD learning is used to estimate the value of a scene as a respawning point.

The rest of this paper is organized as follows. In the next section, the FPS game of this paper is explained. We have to prepare the game as an exercise book in which whoever can learn new knowledge regardless of any skill and knowledge. So in this game zombies are introduced, which everyone already knows they attack a player, but no one knows actual behavior. In the third section, the algorithm recommends a better respawning point for a student to expect to acquire the necessary knowledge and skill is proposed. The final result of the simple experiment by 20 university students is shown.

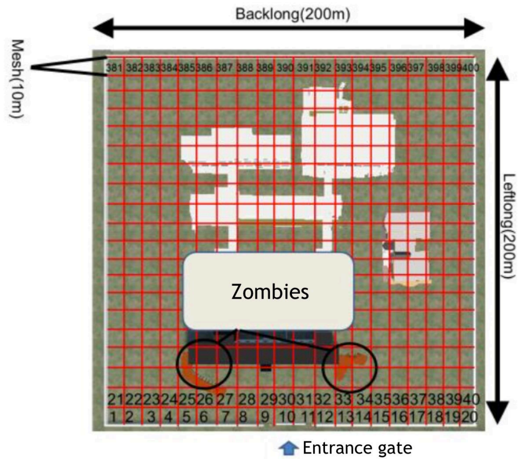


Fig. 2. Mesh Grid of the Field

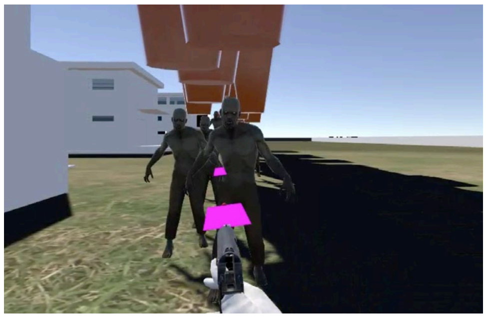


Fig. 3. Game scene

2. The E-learning of FPS style which it can learn how to escape from zombies

2.1. *The overview*

This FPS video game is made by Unity, a 3D game design tool. Fig.1 illustrates the game field. By learning this game a trainee learns skills to escape from zombies. This is the aim of this training and we want to shorten the time to acquire this skill.

Buildings of a high school are simulated in the cyberspace. As Fig.1 shows the school is surrounded by walls and there is only one gate to go outside from the inside of the school. 50 zombies are set around the gate. A trainee becomes an avatar of a high school student, aiming to reach the gate, starting from the furthest point from the gate. Zombies can know the location of the avatar precisely and they go toward the avatar by using Unity's Navigation function which manages object avoidance simultaneously. If a zombie collides the avatar the game trial is over. The positions of all time of the

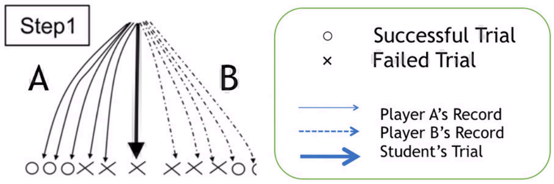


Fig. 4. Step1

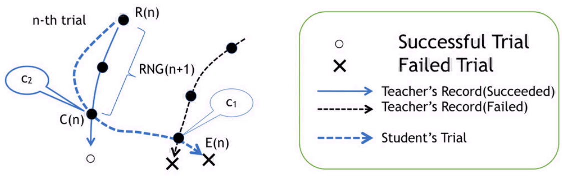


Fig. 5. Step2

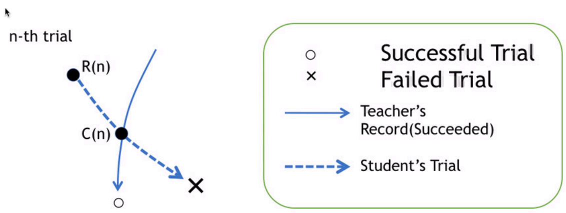


Fig. 6. Step3

zombie and the avatar are recorded using the mesh number of Fig.2.

2.2. *"Respawn" and "respawning point"*

Fig.3 illustrates a scene of this lesson material. The right hand of the student's avatar is shown below the center of this image. The avatar is controlled by the keyboard also its viewpoint is changed by the mouse. When the player reaches the gate, we consider it to acquire the skill and the game is successfully over. On the other hand, the player has to retry the game again when it is caught by zombies. When retrying, call the respawning procedure to change the position of the player and zombies and start the game again. The allocation of the player and zombies is called the respawning point.

3. The proposed respawning algorithm

First, we suppose that successful and unsuccessful attempts have been collected in advance are made into a database. By using the database a respawning point that is easy to learn is generated.

If a certain play follows almost the same trajectory as one of the successful plays and it makes a big deviation from the middle and fails, it suggests that there is at least one operation of the avatar that caused failures somewhere before the deviation.

If we can increase the experience of making the mistake, it will contribute to shortening the acquisition time. To make a rational recommendation even when there is only a relatively small number of times of data, the contribution of each scene in the range where the wrong operation may have been performed is estimated by trial and error. After that, the respawning point is recommended by using this value. In the proposed algorithm, a teacher player is first selected to reduce the amount of calculation required to select the successful play that should be used as a reference.

The respawning point is automatically generated in three stages after a play failure.

(Step1) First, select the teacher player by comparing the record of the player registered in the database and the student's play history. (Step2) Next, select a successful play closest to the filed play from the history of the selected teacher player, and select a promising point as the respawning point from the initial state at the next trial. (Step3) Update the value of the respawning point adopted according to the result of the trial.

The following briefly describes each of these steps.

3.1. *Step1: Selection of teacher player*

Compares the player-level data recorded in the database with the student data, and selects a teacher player. Many methods can be thought of. In our current system, we adopt the maximum information gain approach. The teacher who can expect the largest information gain if the student can behave as it does is selected. The information gain uses KL divergence as follows.

$$KL(P \| Q) = \sum_{i} P(i) \log \frac{P(i)}{Q(i)} \qquad (1)$$

where P(i={success, failure}) is the ratio at which a given player has reached state i in the past from a given state, and Q is the same ratio of the current student. Fig.4 shows an example of this step of the selection of the teacher player, where 5 trial data each of players A and B are found in the database near the student's trials. In that case, therefore P of A (i = success) = 3/5, P of B (i = success) = 2/5, where the probability of A as the teacher of the student is high.

3.2. *Step2: Selection of Teacher's Successful Record*

In Step 2 firstly calculates the core point *C(n)*. The core point *C(n)* is the intersection closest to the state of the student at the end of the n-th play in the set of successful trials of the teacher player. Fig.5 shows an example. Suppose now that the n-th play failed and ended. Here, it is assumed that c_1 and c_2 exist during the teacher's trials near point *E(n)* at the end of the student's play. In this case, c_2 is the core point *C(n)* because the closest intersection c_1 belongs to a failed trial.

3.3. *Step3: Learning of Content Recommendation Skills*

In Step 3 the value of the selected respawning point is updated by using Temporal Difference Learning (TD learning) [4].

$$V\,(R(n)) \leftarrow V\,(R(n)) + \alpha[rn + \gamma V\,(C(n)) - V\,(R(n))] \tag{2}$$

In Fig.6, the value *V(R(n))* of the respawning point *R(n)* is updated on the n-th play. *V(C(n))* is the value of the n-th core point. r_n=1 when the trial is successful, otherwise r_n=0. This *V(・)* will be higher if it contributes to the improvement of students' skills, and it will have a lower value otherwise. By learning this value through trial and error, it can be expected to select a respawning point where learning can be performed more efficiently.

4. Experiment

We spent about two weeks playing the game freely in the laboratory to create the database. Some examples of the trajectories recorded in the database are shown in Fig.7. The solid line and the broken line represent the difference between the players, and the ○ and × marks represent the success case and the failure case. We found that there are two ways to reach the goal without touching the zombies: running through the wall as shown in this figure and passing through the zombies.

The evaluation experiment was conducted on 20 cadet students from February 20 to 22, 2019. The flow of the experiment is described in Fig.8. The 20 people were divided into three groups: 1) a group using teaching materials without a recommendation function, 2) a group recommending respawning points using the proposed method, and 3) a group restarting from a failed point. Group 1 always starts the game from the beginning if it fails. The self-study time was 10 minutes for each student. After that, four patterns of games where the zombies were located were different, and the learning effect of each group was evaluated. The four patterns used are shown in Fig.9. Pattern 0 was used during the self-study, and patterns 1 to 3 were newly prepared. Pattern 1 is a case where zombies are gathering near the goal, pattern 2 is a case where zombies are behind a building, and pattern 3 is when the zombies are in the immediate vicinity. Students will be tested for their understanding of the layout of school buildings and how zombies move.

Zombies of Pattern 0 appears at almost the same position as during self-study, so you can check the level of understanding of the learning content. Looking at the results of pattern 0, the average number of plays in Group 1 is 2.50, which is about twice the 1.17 times of the proposed method. Also, Group 2 to propose is shorter than 1.50 of Group 3 to start over from the point of failure. From this, it can be seen that the proposed method can learn more lean and stable skills.

In Patterns 1 to 3, Group 2 of the proposed method reached the goal with an average of 1.84 times, which was far quicker than the other two groups (2.37 times in Group 1, 2.34 times of Group 3).

Focusing on the highlights, it was found that the more experience it encounters the zombies, the fewer trials they have. That is, in the case of pattern 1, Group 1 is the best, in the case of pattern 2, Group 2 is the best, and in the case of pattern 3, Group 3 is the best. Also, the playtime of successful trials is the shortest on average for Group 2 at 28.45 seconds, and it can be seen that for any pattern, Group 2 of the proposed method achieves lean movement.

From the above, it was found that by using the proposed method, highly versatile skills can be acquired by self-study.

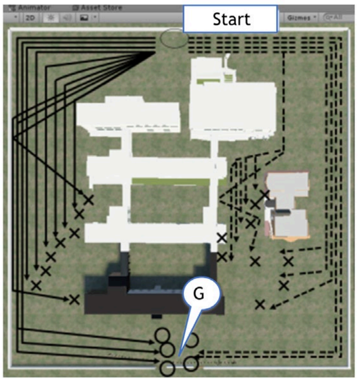


Fig. 7. Examples of Trajectories of plays

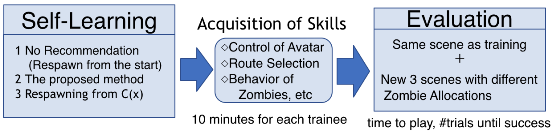


Fig. 8. Experiment Flow

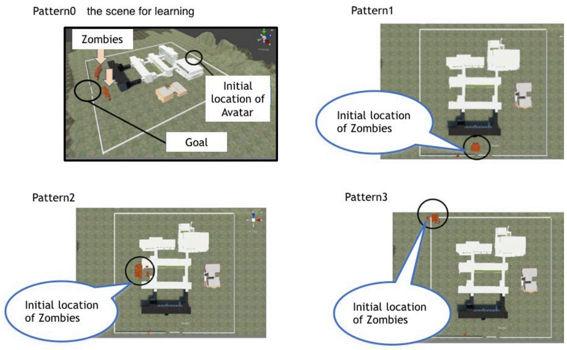


Fig. 9. Scenes for the evaluation

1 No Recommendation	8 trainees				
pattern	0	1	2	3	average
#trials to success	2.50	1.83	3.83	1.33	2.37
ave. time to goal	33.30	32.53	34.82	32.25	33.22
2 The proposed method	6 trainees				
pattern	0	1	2	3	average
#trials to success	1.17	1.83	2.67	1.67	1.84
ave. time to goal	28.26	28.40	30.07	27.06	28.45
3 Respawn from C(x)	6 trainees				
pattern	0	1	2	3	average
#trials to success	1.50	1.17	4.67	2.00	2.34
ave. time to goal	28.56	28.97	28.39	28.93	28.46

Fig. 10 Results

5. Conclusion

This paper proposes recommending respawning points according to the trainee's learning status, to reduce the trainee's time to acquire skills with real-time video game type E-learning teaching materials. The proposed method estimates the value of each scene as a start point where the task can be expected to be solved quickly. The proposed method was evaluated using 20 test subjects, and it was confirmed that the proposed method was able to learn more efficiently and comprehensively than the no recommended method and the method of restarting from a failed point.

References

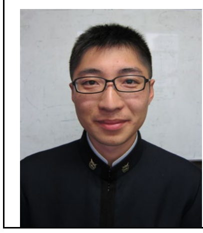
1. Geoffrey A. Frank, Robert F. Helms Ii, David Voor, “Determining the Right Mix of Live, Virtual, and Constructive Training”, Proceedings of the Interservice/Industry Training, Simulation and Education Conference. Orlando, FL National Training Systems Association, pp27-30, 2000
2. Georgios N. Yannakakis, Julian Togelius, "Artificial Intelligence and Games", Springer, 2018
3. Kevin D. Saunders, Jeannie Novak, "Game Development Essentials: Game Interface Design 2nd edition", Delmar cengage learning Pub, 2013
4. R. S. Sutton, A. G. Barto, “Reinforcement Learning: An Introduction,” A Bradford Book, 1998.

Authors Introduction

Dr. Masao Kubo

Masao Kubo is an Associate Professor of Department of Computer Science at National Defense Academy in Japan

Mr. Takashi Ueno



Mr. Takashi Ueno graduated from Department of Computer Science at National Defense Academy in Japan, 2019. He belongs to Japan Maritime Self-Defense Force.

Dr. Hiroshi Sato

Hiroshi Sato is an Associate Professor of Department of Computer Science at National Defense Academy in Japan.