Video Game Agents with Human-like Behavior using the Deep Q-Network and Biological Constraints

Takahiro Morita[1] and Hiroshi Hosobe[2]
[1]Graduate School of Computer and Information Sciences, Hosei University, Tokyo, Japan
[2]Faculty of Computer and Information Sciences, Hosei University, Tokyo, Japan

Keywords: Machine Learning, Deep Q-Network, Biological Constraint, Video Game Agent.

Abstract: Video game agents that surpass humans always select the optimal behavior, which may make them look mechanical and uninteresting to human players and audience. Since score-oriented game agents have been almost achieved, a next goal should be to entertain human players and audience by realizing agents that reproduce human-like behavior. A previous method implemented such game agents by introducing biological constraints into Q-learning and A* search. In this paper, we propose video game agents with more entertaining and more practical human-like behavior by applying biological constraints into the deep Q-network (DQN). Especially, to reduce the problem of the conspicuous mechanical behavior found in the previous method, we propose an additional biological constraint "confusion". We implemented our method in the video game "Infinite Mario Bros." and conducted a subjective evaluation. The results indicated that the agents implemented with our method were rated more human-like than those implemented with the previous method.

1 INTRODUCTION

In recent years, research on various game agents has been actively conducted. In 2017, Ponanza, AI of Shogi (Japanese chess), won a professional player. In the same year, AlphaGo (Silver, et al., 2016), AI of another board game Go, also won professional players. In addition, agents for incomplete board games such as The Werewolves of Millers Hollow (Katagami, et al., 2018) and for video games such as Tetris and Space Invaders have been widely studied in the field of AI. For example, general-purpose AI using deep Q-learning was created (Mnih V. , et al., 2013); it automatically generated features and evaluation functions for simple video games such as Pong and Breakout by using only game screens.

Game agents that surpassed humans always select the optimal behavior, which may make them look mechanical and uninteresting to human players and audience. Since score-oriented game agents have been almost achieved, a next goal should be to entertain human players and audience by realizing game agents that reproduce human-like behavior. Recent video games are often focused on entertaining players by providing game agents called "non-player characters" that behave like humans. Although various studies have been conducted, most of them need to define or learn human-like behavior, which puts heavy workloads on game AI developers.

To solve this problem, a previous study (Fujii, Sato, Wakama, Kazai, & Katayose, 2013) proposed a method to realize human-like behavior based on biological constraints, i.e., "fluctuation", "delay", and "tiredness". Their method reduced development cost and could be introduced into various game genres. Since it used Q-learning and A* search, which are simple methods for machine learning, the skill of their video game agent was as low as a human beginner or an intermediate-level player.

In this paper, we propose video game agents with more entertaining and more practical human-like behavior by introducing biological constraints into the deep Q-network (DQN) (Mnih V. , et al., 2015), which is a machine learning method often used for video game agents. In addition, we propose an additional biological constraint "confusion", which can occur when the amount of information exceeds what humans can process.

We conducted two experiments. First, we performed a preliminary experiment to examine how the implementation with DQN make a difference. For this experiment, we prepared Q-learning and DQN with and without Fujii et al.'s biological constraints,

Morita, T. and Hosobe, H.
Video Game Agents with Human-like Behavior using the Deep Q-Network and Biological Constraints.
DOI: 10.5220/0011697500003393
In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 3*, pages 525-531
ISBN: 978-989-758-623-1; ISSN: 2184-433X

applied them to the video game "Infinite Mario Bros.", and subjectively evaluated whether they showed human-like behavior. As a result, the game agent introducing the biological constraints to DQN showed human-like behavior to some extent, but also showed conspicuous mechanical behavior. Second, we implemented the game agent that further applied our new biological constraint "confusion" and conducted subjective evaluation by several participants. As a result, it was shown that our method with the additional "confusion" constraint exhibited more human-like behavior than the methods without the additional constraint.

2 INFINITE MARIO BROS.

Infinite Mario Bros. (Figure 1) is a game that imitates Super Mario World, a 2D side-scrolling action game released by Nintendo. Infinite Mario Bros. infinitely generates stages by using the seed value of pseudo-random numbers given in advance. Much research has been done on video game agents for Infinite Mario Bros.; especially, at the 2009 Conference on Computational Intelligence and Games (CIG) (Togelius, Karakovskiy, & Baumgarten, 2010), a tournament called Mario AI Competition was held for this game. This tournament provided a program for the implementation of game agents. It gives information such as the positions of enemies and other objects around Mario, which allows the easy implementation of reinforcement learning. In this paper, we use this program to implement our game agents.

Figure 1: Infinite Mario Bros.

3 RELATED WORK

There has been research on game agents with human-like behavior. Some studies acquired human-like behavior by imitating human players. A video game agent (Polceanu, 2013) for a first-person shooter called Unreal Tournament 2004 realized human-like behavior by recording the movements of other players in real time and partially reproducing the data. Another video game agent (Ortega, Shaker, Togelius, & Yannakakis, 2013) was focused on the path taken by human players in Infinite Mario Bros. To imitate human players, the agent implemented three indirect methods, i.e., hand coding, a direct method based on supervised learning, and an indirect method based on maximizing similarity measures. As a result, the indirect method based on maximizing similarity measures was the best. However, the game agent required a large amount of data of actual human players for acquiring human-like behavior, which would increase development cost.

To solve the problem of development cost, research on game agents that automatically acquires human-like behavior has been conducted. "Entertainment Go AI" (Ikeda & Viennot, 2013) was realized by enabling a fully-enhanced Go game agent intentionally make human-like mistakes for human players to enjoy playing Go with it. However, with this method, the game AI developer needs to define human-like mistakes, which increases the workload in development. By contrast, biological constraints (Fujii, Sato, Wakama, Kazai, & Katayose, 2013), i.e., "fluctuation", "delay", and "tiredness", which can occur when humans are playing video games, automatically acquired human-like behavior in video game agents by using Q-learning and A^* search. They were applied to Infinite Mario Bros., and the movement of the resulting game agents was subjectively evaluated.

4 PRELIMINARIES

4.1 Q-Learning

Q-learning (Watkins & Dayan, 1992) learns by updating the value estimation function $Q(s,a)$, which is an index of what kind of action a the agent should take under a certain state s. The agent receives a reward according to the result of the action, and updates $Q(s,a)$ using the reward. The update equation is the following:

$$Q(s,a) = (1-\alpha)\, Q(s,a)$$

$$+\alpha[R(s,a,s') + \gamma \max Q(s',a')]$$

where α is a learning rate, and γ is a discount rate. The term $1-\alpha$ represents the current Q value, and the term α represents the value used in learning. This equation causes a higher reward to make a higher updated Q value. In addition, the action with the highest Q value in the current state is selected.

4.2 Deep Q-Network

The deep Q-network (DQN) (Mnih V. , et al., 2015) introduces the concept of a neural network to Q-learning. It is difficult for classical Q-learning to represent continuous states because the number of states becomes enormous. By contrast, it is possible for DQN to define a complicated state by using a neural network to estimate the Q value and obtaining an approximate function of the Q value. Specifically, the action with the highest Q value will be known if the agent obtains the optimal action function by an approximate function using a neural network and estimates the Q value for each action in a certain state. $Q(s_t, a_t)$ is calculated by the Q-network (Figure 2), which is a neural network whose input layer is a certain state s_t and whose output layer is an action a_t.

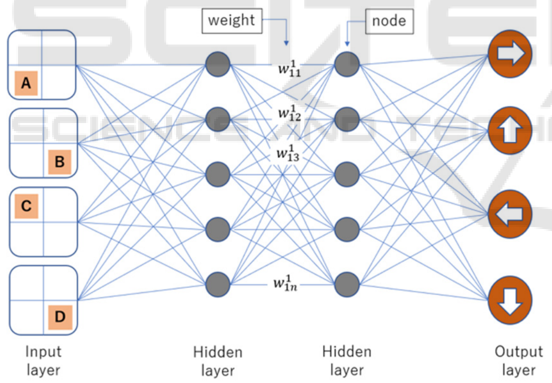

Figure 2: Deep Q-Network.

4.3 Biological Constraints

The video game agent that learns autonomously often plays too accurately. More specifically, the reaction speed to enemies and fields is too fast, constant actions are always repeated accurately, and movements such as walking and jumping are often being constant and unnatural. To prevent such problems, biological constraints (Fujii et al., 2013) that are "physical constraints" and "desire to survive" were combined with a machine learning method that performed autonomous learning. The biological constraints make it possible to eliminate the above-mentioned behavior in a video game agent and eliminate the "intentionality" and associated "boringness" that can occur in adjustment methods such as making mistakes and weakening. They defined and introduced the following biological constraints in Q-learning:

- "Fluctuation" in the sensor system and the motion system. Since human players cannot always accurately judge what they see and also they cannot always accurately operate it, they often make misunderstandings and mistakes in operations. To reproduce this, Gaussian noise was added to the information observed by the game agent.
- "Delay" from perception to motor control. Human players have delay from the observing information to operating. To reproduce this, the game agent observed information a few frame ago.
- "Tiredness" of key operation. Human players get tired from operating the game controller and drop their performance. Therefore, a negative reward was given each time the key operation was changed.

5 PROPOSED METHOD

In our research, as an extension of the research of Fujii et al., we propose a method that introduces biological constraints into DQN, which is often used in the implementation of game agents. This method has three main goals:

- Enable the automatic acquisition of human-like behavior;
- Enable the development of a game agent with an optimal balance between human-like behavior and game skill;
- Automatically acquire sufficiently human-like behavior even if it is applied to various game genres.

The introduction of biological constraints to DQN is basically the same as that the previous method implemented in Q-learning. "Fluctuation" is reproduced by adding Gaussian noise to the image put in the input layer. "Delay" is reproduced by using the state s_{t-n} which is n-frame before the current frame. "Tiredness" is reproduced by obtaining a constant negative reward after changing the action of the operation key.

However, we found a serious problem from the results of the preliminary experiment described in Section 6. The game agent introducing DQN and only these three biological constraints performed the optimal behavior that exceeded the skills of humans, which made it difficult to obtain human-like behavior.

To solve the problem, we propose a new biological constraint “confusion”. When a human player obtains too much information, it becomes difficult for the human player to operate the controller as usual. We consider that “confusion” can suppress the mechanical behavior of a game agent, such as easily defeating enemies even if it is surrounded by many enemies. We reproduce “confusion” by increasing “fluctuation” and “delay” parameters of biological constraints when a certain number of objects including enemies are displayed around the player controlled by the agent.

6 PRELIMINARY EXPERIMENT

As a preliminary experiment in our research, we applied Q-learning and DQN with and without the biological constraints of “fluctuation”, “delay”, and “tiredness” to Infinite Mario Bros, and compared their behaviors. In this preliminary experiment, we by ourselves observed the behaviors of the learned video game agents.

In performing the subjective evaluation, we focused on the following points:

- Does Mario move like a human player?
- Does the movement surpass human reflexes?
- Does the agent hesitate in front of an enemy or hole, or try to jump over it safely?
- Is there inconsistency between jump height and dash speed?

The result of the preliminary experiment is summarized in Table 1. First, we compared game agents without biological constraints. From Q-learning without biological constraints, we did not observe any behavior that felt particularly human-like. DQN without biological constraints showed more mechanical movements than Q-learning, and behaved accurately and optimally for enemies and holes without any hesitation. In addition, it seemed that the dash speed did not change because it progressed almost without stopping.

Next, we compared game agents with and without biological constraints. In both Q-learning and DQN, the introduction of the biological constraints changed the movements from the one that could cause a dodge on the verge to the one that caused safe jumps with a little surplus in front of an enemy or a hole. It seemed that the reaction speed was suppressed by biological constraints. Also, when there were many enemies, game agents were waiting until it was possible to advance, instead of forcibly breaking through by accurately operating the controller. From these results, it is thought that human-like behavior was acquired to some extent.

Finally, we compared Q-learning and DQN which introduced the biological constraints. We were able to confirm the human-like behavior of Q-learning, such as hesitation in front of enemies and holes, and stopping to emphasize safety. However, its final score was not high, and it looked like a beginner’s play. By contrast, the final score of DQN was sufficiently high, and we were able to confirm its human-like behavior to some extent. However, it was observed that there was much movement due to the optimal behavior, and its mechanical movement seemed conspicuous.

Table 1: Result of the preliminary experiment.

Machine learning	Biological constraints	Subjective evaluation
Q-learning	None	No human-like movement; Intermediate player
Q-learning	Imposed	Hesitation in front of enemies, holes, etc.; Intermediate player
DQN	None	Mechanical movement; Advanced player
DQN	Imposed	Slightly human-like movement; Advanced player

7 MAIN EXPERIMENT

7.1 Procedure

In the main experiment, we implemented a game agent without biological constraints, a game agent with biological constraints (with 4 frames of delay), and a game agent with additional biological constraint “confusion” (normally with 4 frames of delay, and during confusion with 6 frames of delay and an increased fluctuation) by using Q-learning and DQN. In these game agents, coins, blocks, and items were ignored. Each learning trial had 50,000 games, the learning rate α was 0.2, the reduction rate γ was 0.9, and the random selection probability for the ε-greedy method was 0.07.

We performed subjective evaluation on human-like behavior and playing skills. We recruited 10 participants ranging from 19 to 23 years old, who had played the Super Mario series more than 10 hours in total. This subjective experiment was conducted online.

Table 2: Prepared videos.

Label	Human and NPC	Biological constraints	Play time
[Q, None]	Q-learning	None	20 sec
[Q, Imp]	Q-learning	Imposed	22 sec
[Q, Imp, Con]	Q-learning	Imposed (+ confusion)	23 sec
[DQN, None]	Deep Q-Network	None	20 sec
[DQN, Imp]	Deep Q-Network	Imposed	22 sec
[DQN, Imp. Con]	Deep Q-Network	Imposed (+ confusion)	18 sec
[Player]	-	-	25 sec

Table 2 summarizes seven videos that we prepared, where game agents or a human player were playing Infinite Mario Bros. Three videos were on game agents using Q-learning, other three videos were on game agents using DQN, and the other video was on a human player. Each video was trimmed around 20 seconds where holes and enemies appear. To confirm the differences between the actions of game agents with or without confusion, we prepared scenes that Mario was surrounded by enemies.

The procedure of this subjective experiment was as follows. First, we showed the screen of playing Infinite Mario Bros. for about 1 minute and used it as a reference to evaluate how clearly the videos showed human-like behavior. Next, participants were showed randomly selected seven videos, and asked to answer two questions on a seven-point scale: “Is this player human-like?” (1 is the most mechanical movement and 7 is the most human-like movement), “How much game skill does this player have” (1 is a beginner and 7 is advanced). In addition, they were asked to write free comments about each play: for example, “I felt that player hesitated in front of a hole.”

7.2 Result

The result of the experiment is shown in Table 3 and Table 4. First, we compare game agents with Q-learning and DQN. In comparison of [Q, None] and [DQN, None], the results of 2.3 and 1.7 were negative in the question of whether they were human-like. In comparison of [Q, Imp] and [DQN, Imp], there were relatively more positive evaluations; human-like movements were seen in [Q, Imp], which was evaluated 4.0, while [DQN, Imp] was evaluated as more negative than [Q, Imp] such as “dodging enemies and blocks too much”. However, the playing skill of [Q, Imp] was not evaluated high. In comparison of [Q, Imp, Con] and [DQN, Imp, Con], [DQN, Imp, Con] was evaluated higher than [Q, Imp, Con] in terms of human-like behavior. Furthermore, in terms of playing skill, [DQN, Imp, Con] was evaluated higher than any Q-learning game agents. Therefore, it can be seen that [DQN, Imp] was able to acquire human-like movements while having the skill to play well.

Next, we compare Q-learning and DQN. In Q-learning, [Q, Imp] was able to get more human-like behavior than [Q, None], but compared with [Q, Imp], [Q, Imp, Con] gave negative results, and there were comments such as “It felt unnatural”. By contrast, in DQN, since biological constraints were increased, the evaluations of human-like movements were increased to 1.7, 3.3 and 4.3.

Finally, we compare [Player] and [DQN, Imp. Con]. In the evaluation of human-like behavior, [DQN, Imp. Con] was 0.9 higher than [Player] in terms of play skill. Conversely, [DQN, Imp. Con] was 1.1 lower in the evaluation of human-like behavior. Therefore, we can observe that there is still a gap between [Player] and [DQN, Imp. Con]. In addition, in [Player], there was a positive opinion that “There was a behavior where Mario seemed to stop and think about the next move”. By contrast, in [DQN, Imp, Con], there were negative opinions such as “It was a little unnatural because Mario continued to act without stopping too often.”

Table 3: Result of the main experiment. Each number shows an average of the participant’s answers (1 is the most negative and 7 is the most positive).

Label	Human-like	Player skill
[Q, None]	2.3	3.6
[Q, Imp]	4.0	3.3
[Q, Imp, Con]	3.6	2.5
[DQN, None]	1.7	5.9
[DQN, Imp]	3.3	5.0
[DQN, Imp, Con]	4.3	4.7
[Player]	5.4	3.8

8 DISCUSSION

In this section, we discuss four things based on the results of the main experiment. First, we discuss the impact of implementing confusion in game agents. In the result of the main experiment, [DQN, Imp, Con] was able to acquire more human-like behavior than [DQN, Imp] in contrast to little declining playing skill. It was shown that [DQN, Imp, Con] was able to obtain more human-like behavior than the method of the previous research. In addition, in the free comments about [DQN, None] and [DQN, Imp], there were many negative opinions about human-like behavior such as easily defeating enemies in situations where Mario was surrounded by enemies. By contrast, In the free comments about [DQN, Imp, Con], there was no such opinion. We consider that our method solved the problem of a game agent that unnatural mechanical behavior occurred in scenes where Mario was surrounded by enemies and blocks.

Next, we discuss whether the biological constraints introduced into DQN were sufficient. In the result, there was still a gap in evaluation between [DQN, Imp, Con] and [Player]. Since DQN without biological constraints has a high playing skill, we consider that even if we add more biological constraints, we can acquire human-like behavior without lowering playing skill much. In our proposed method, “confusion” solves the problem on the condition such as when the enemy is dense. There are still many biological constraints on such specific conditions. We consider that implementing them as much as possible will solve the mechanical behavior problem and the game agent can look like an almost human play.

We discuss in what situations the game agent obtained by this proposed method would be useful. In recent years, game agents for NPCs for competitive games have been trained by using a large amount of battle data, but the cost of development is huge. By using our proposed method, it is possible to automatically acquire a game agent that shows human-like behavior and that players can enjoy. As for other usage, in recent years, there are many people who enjoy watching game play videos. It is thought that it is possible to create a game agent that is fun for those people to watch as entertainment.

Finally, we discuss the experimental method. In our paper, human-likeness was verified by subjective evaluation by participants. However, in such a subjective evaluation experiment, the result may vary greatly depending on participants because the participants’ feeling is different individually, and the persuasiveness of the results may be relatively low. To solve this problem, it is necessary to quantitatively define human-likeness and conduct experiments. However, human-likeness is ambiguous and difficult to measure quantitatively. Almost all previous studies only employ subjective evaluation experiments like our research. If human-likeness is defined quantitatively, the ambiguity of evaluation will disappear, and it will make a great contribution in the field of human-like agent generation.

9 CONCLUSIONS AND FUTURE WORK

In this paper, we combined DQN with biological constraints to realize a game agent with human-like behavior. The purpose of our research is to automatically generate a game agent that have a game skill of an advanced player and perform human-like behavior. From the result of the preliminary experiment, the DQN-based agent obtained more mechanical behavior than the agent based on Q-learning when only the biological constraints “fluctuation”, “delay”, and “tired” were introduced. To solve the problem, we proposed a new biological constraint “confusion”. As a result of the subjective evaluation experiment, it was shown that the game agent that introduced our proposed method performed more human-like behavior than the method of previous research.

As our future work, we need to consider two points:

- Explore new biological constraints;
- Apply our method to other game genres.

Since a game agent with biological constraints learns to perform the optimal actions within those constraints, we consider that more complete human-like behavior can be acquired with high game skills by introducing additional biological constraints. An example of a biological constraint that we are considering is “carelessness”. It is a psychological constraint that makes people act without thinking deeply when things are going smoothly. We should verify whether human-like behavior can be obtained by adding such new biological constraints.

In addition, one of the purposes of this proposed method is that it can be applied to any game genres, like the research by Fujii et al. Therefore, it is necessary to apply it to games other than Infinite Mario Bros. and verify whether it can perform human-like behavior sufficiently. However, the problem is that, in games with little movement such as simulation games, it is not possible to visualize sequential actions as in action games or shooting games. This makes it extremely difficult to judge whether the actions are human-like, which requires further investigation.

REFERENCES

Fujii, N., Sato, Y., Wakama, H., Kazai, K., & Katayose, H. (2013). Evaluating Human-like Behaviors of Video-Game Agents Autonomously Acquired with Biological Constraints. *Proc. ACE, LNCS, 8253*, 61-76.

Ikeda, K., & Viennot, S. (2013). Production of Various Strategies and Position Control for Monte-Carlo Go—Entertaining Human Players. *Proc. IEEE CIG*, 1-8.

Katagami, D., Toriumi, H., Osawa, H., Kano, Y., Inaba, M., & Otsuki, T. (2018). Introduction to Research on Artificial Intelligence Based Werewolf (in Japanese). *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, 30*(5), 236-244.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing Atari with Deep Reinforcement Learning. *arXiv, 1312.5602*.

Mnih, V., Kavukcuoglu, K., Silver, D., Ruse, A. A., Veness, J., Bellemare, M. G., & Graves, A. (2015). Human-Level Control through Deep Reinforcement Learning. *Nature, 518*, 529-533.

Ortega, J., Shaker, N., Togelius, J., & Yannakakis, G. N. (2013). Imitating Human Playing Styles in Super Mario Bros. *Entertainment Computing, 4*(2), 93-104.

Polceanu, M. (2013). Mirrorbot: Using Human-Inspires Mirroring Behavior to Pass a Turing Test. *Proc. IEEE CIG*, 1-8.

Silver, D., Aja, H., Maddison, C. J., Guez, A., Sifre, L., Driessche, G. v., & Schrittwieser, J. (2016). Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature, 529*, 484-489.

Togelius, J., Karakovskiy, S., & Baumgarten, R. (2010). The 2009 Mario AI Competition. *Proc. IEEE CEC*, 1-8.

Watkins, C. J., & Dayan, P. (1992). Q-Learning. *Machine Learning, 8*, 279-292.

APPENDIX

Table 4: Free comments of the participants about human-like behavior.

Label	Positive evaluation	Negative evaluation
[Q, None]	▪ Not jump a lot	▪ Dodge a hole with a minimum jump ▪ Go forward nonstop by defeating enemies ▪ Feel unnatural ▪ Defeat enemies easily
[Q, Imp]	▪ Stop before a hole or an enemy ▪ Play safely ▪ Not make constant movement	▪ Make improper jump timing ▪ Feel unnatural
[Q, Imp, Con]	▪ Stop before a hole or an enemy ▪ Play safely ▪ Not make constant movement ▪ Try to avoid many enemies	▪ Feel unnatural when the player was surrounded by enemies
[DQN, None]	▪ Play like an advanced human player	▪ Dodge a hole with a minimum jump ▪ Go forward nonstop by defeating enemies ▪ Move faster than a human player ▪ Defeat enemies too easily ▪ Feel unnatural when the player was surrounded by enemies
[DQN, Imp]	▪ Play like an advanced human player ▪ Not make constant movement ▪ Stop before a hole or an enemy ▪ Play safely	▪ Make improper jump timing ▪ Feel unnatural ▪ Go forward nonstop by defeating enemies ▪ Defeat enemies too easily ▪ Feel unnatural when the player was surrounded by enemies
[DQN, Imp. Con]	▪ Play like an advanced human player ▪ Not make constant movement ▪ Stop before a hole or an enemy ▪ Play safely	▪ Feel unnatural ▪ It was a little unnatural because it continued to act without stopping too much
[Player]	▪ Stop before a hole or an enemy ▪ Play safely ▪ Not make constant movement ▪ There was a behavior where Mario seemed to stop and think about the next move	▪ Feel unnatural ▪ Make improper jump timing