Reinforcement Learning for Improving Flappy Bird Game

Shiyao Wei *

The senior high school attached to the ShanDong normal University, Jinan, China

* Corresponding Author Email: 100561@yzpc.edu.cn

Abstract. Currently, Artificial Intelligence becomes popularity among the human daily life, like games, Internet and so on. Authors has shown that in the game filed the Artificial Intelligence always have better performance than human beings, so in this article, the author wants to use AI to carry out on an old- fashion fame called Flappy Bird. This study aims to determine the specific method why AI has better performance than human beings. In this context, the author based on the process of experiments: It mainly used reinforcement learning model (Acting based on feedback from the environment, through continuous interaction with the environment, trial, and error, to ultimately accomplish a specific purpose or to maximize the overall benefits of the action) and supervised learning model (the process of making the machine learn a large amount of sample data with labels, training a model and making the model get the corresponding output according to the input) to improve the Flappy Bird and both two method are belonging to the machine learning. In addition, this study alters the layout of the game, including pipe, appearance of agent, and background of the game in order to make a more fashionable game. Furthermore, this study increases the number of agents, which makes it easier for agent to achieve higher score. Last but not the least, author establish a archive point, which means if the player face operation mistake and lead to game over, they bird will relive before passing the last pipe.

Keywords: Reinforcement learning, Machine Learning, Q-learning, Flappy Bird.

1. Introduction

Reinforcement Learning (RL) is the science of decision making. It is about learning the optimal behavior in an given situation to maximize the reward. This optimal behavior is learned through interactions with the environment and observations of how it responds, similar to children exploring the world around them and learning the actions that help them achieve a goal [1]. Furthermore, the reinforcement learning has been extensively used in our daily life. Take an example of AlphaGo, when playing Go, it is not realistic to consider every possibility and program it, so a high dimensional way to solve this problem is needed: it should make decisions autonomously under uncertainty occasions in order to reach the final target.

Flappy Bird is a mobile game developed by Vietnamese video game artist and programmer Dong Nguyen under his game development company [2]. The game is a side-scroller where the player controls a bird, attempting to fly between columns of green pipes without hitting them [3]. In 2004 Flappy bird became a popular game with the public. After that, the public found that it can be achieved in the Pygame but more importantly it lends itself well to reinforcement learning. The goal of the game is to keep the bird alive as long as possible to achieve the highest score. Due to the neutral effect, only the player who does not click the bird will fall to the ground, the game will be technical. In addition, the bird needs to balance itself through the pipe through the player's constant clicking. The pipes also limit the height at which the birds can fly. They must be kept at a certain height as they pass through the pipes. Birds can also die if they hit the pipes. In this case, the project focus on how to use reinforcement learning to improve the Flappy bird in order to achieve the higher score.

In the early field of study, Google Deepmind Minih et al. had used reinforcement study on Atari 2600, which achieves a huge success in the world [4]. To be more specific, the Deep Q Network (DQN) was used to evaluate the Q function of Q-learning, and also use experience replay to de-correlate experience. It illustrated how AI can learn to run a game without knowing anything about it. In order to ensure the accuracy, the research applied same module on 7 Atati games and the result is impressive: AI performed better than the professional player in all 7 games. In addition, Algha Go

is also an impressive Artificial Intelligence, which published by Demis Hassabis in Google company. Their team based on supervised leaning and reinforcement learning to design a reward function for Alpha Go (0 reward for the middle moments, 1 reward for victory and -1 reward for failure at the end moments), a value function to optimize the goal, and a Monte Carlo number search algorithm to simulate the game.

Although this practice can imply on the games, further improvements involve prioritizing experience replay, more efficient training, and better stability when training, which makes games easier to operate. To be more specific, this study tries to reduce the loss to +1 or -1, in order to make the agent learn the desired skill more quickly and reliably. And this study set a value C for DNQ to update every C instead of updating the target network every iteration.

In this study, 2 convolutional layers and 2 linear layers were applied, using reward function, and a Monte Carol number search algorithm. agent approximately 10-15 hours was trained using the Q-learning and supervised learning and reinforcement learning was combined for allowing this study to get a much better result that this study get an average reward of 16.1 and 3.5 for the network. In addition, number of agent and resume game from death was tried to increase.

2. The basic fundamental method

2.1. Flappy bird

The goal of the game is to keep the bird alive as long as possible to achieve the highest score. Due to the neutral effect, only the player who does not click the bird will fall to the ground, the game will be technical. In addition, the bird needs to balance itself through the pipe through the player's constant clicking. The pipes also limit the height at which the birds can fly. They must be kept at a certain height as they pass through the pipes. Birds can also die if they hit the pipes.

Flappy Bird is a famous mobile game in which the player has to guide a bird through the gaps between regularly disposed of pipes. The gameplay is very simple: at every instant, the player can choose between two actions: doing nothing or flap thus letting the bird descend or tapping the screen, thus making the bird fly upward. The general setup of the game (can be seen in Figure 1.) And the pipeline through which the agent passes is the only environment variable. So, this is a very good problem for implementing reinforcement learning to improve things.

So, learning a control policy directly from high-dimensional is necessary to carry out. Image inputs in environments that don’t provide rewards frequently is a long-standing problem in reinforcement learning, that has crucial implications for robotics and autonomous vehicles. Reinforcement learning module is used into Flappy Bird that learns control policies directly from image observations and from feedback received when the bird hurts an obstacle. Generic algorithm is carried out that looks at the raw pixel values of the game frames and trains the agent to take the right decision at any time during the game. Teaching Flappy Bird agent how to fly and go through obstacles using a variant of Q-Learning. Investigating the impact of image preprocessing and other ways to improve training time and rewards achieved by the agent.

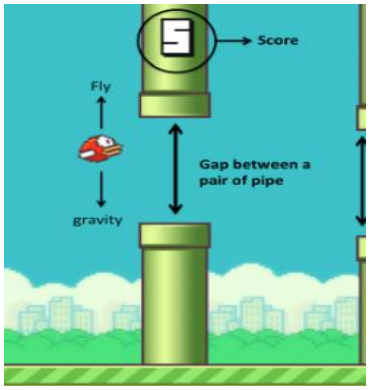

Figure 1. The schematic diagram of the Flappy bird.

2.2. The method carrying out in Flappy Bird

Q-Learning is a Reinforcement learning policy that will find the next best action, given a current state. It chooses this action at random and aims to maximize the reward and it is a model-free, off-policy reinforcement learning that will find the best course of action, given the current state of the agent. Depending on where the agent is in the environment, it will decide the next action to be taken.

In machine learning, training a model often requires a large amount of data [5-10]. The more complex the model, the more data it needs. However, our final data may not be 100% reliable. It may have wrong or missing values.

Reinforcement learning solves this problem almost perfectly because it does not require so much data. Reinforcement learning is about letting a model make its own next decision and then keep trying until it finds the best result. In the flappybird, the Q-learning algorithm is used in reinforcement learning to continuously train the bird, and finally achieve the ending of "eternal life". Regarding Q-learning, it starts with its transfer rules.

$$\mathrm{Q(s,a)} = \mathrm{R(s,a)} + \gamma . \underset{\tilde{a}}{max}\{Q(\tilde{s},\ \tilde{a})\} \tag{1}$$

First, creating an initial Q-Table with all values are 0. Then select an action for the current state and execute it according to the current Q-Table. In order to maintain a stable high survival rate for the bird, set the reward of alive to 0 and the reward of death to -1000. After taking the action and getting the reward, Q function is used to update Q(s,a). Therefore, three things are used to improve Flappy Bird. Because training time is too long. It takes 10-15 hours to train the agent using normal Q-learning

3. Improvements in Flappy Bird

3.1. Combining the supervised learning and reinforcement learning

The finding is that in the first ten minutes of training when the score was around three points, the bird would sometimes do nothing when facing the tube, which means it would just bump into it.(show in the Figure 2) Although this problem will be gradually improved because Q-learning will give the agent a punishment to force him to change. When the agent passes through the first tube, he will get a reward, but only by constantly updating the reward can the agent know that the only right thing is to pass the tube as it gives it a reward. But if the agent chooses to Do Nothing when it's facing the tube, or if the agent chooses to click but it's too late, the Agent crashed into the bottom of the tube eventually. So, if the agent does some action and leads to a crash on the tube, the game should label this action as “bad”; more precisely, setting the max score for 1 and let it play for 10 times and observing the agent reaction can be achieved. In addition, action which crashed on the tube and labels them as “bad” action, is picked out. And label the action that leads the agent successfully enter the tube as “good”, no matter if it finally crosses the tube. So as a result, the agent will know that their objective is to enter the tube when facing the tubes, and as for how to cross the tube that’s the job of reinforcement learning.

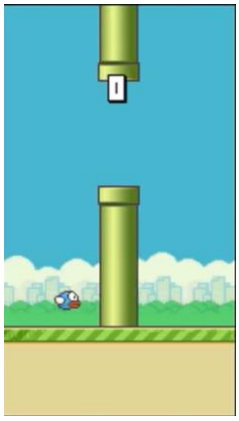

Figure 2. Occasion when bird face the tube

3.2. Increase the number of agents

Author thinks the two agents are considerable. First of all, the two agents can be displayed on the screen, and the three agents may not be displayed together. Second, the two agents should use the same Q-value table, that is to say, they can know what each other has done. For example, the first agent has crashed, but the second agent receives the return value of the first agent, changes its action at the place where the first agent crashed, and survives successfully and this practice can save a lot of time.

But when the agent double there is a problem that it cannot get rid of which is credit assignment. One of the agents may become the lazy agent which means it relies on the policies the other agent learns and stop exploration as the global reward is already very high or it can already survive using the other one's policy. When this happens, it is impossible for them to cooperate so they cannot learn from the other's mistakes. To overcome this issue this article uses Counterfactual Multi-Agent Policy Gradients (COMA) In order to solve the problem of poor cooperation ability between agents, COMA uses critics to replace rewards, and trains one critic as the global critic in a centralized manner.

Critical is only used in the learning process. It can be trained based on all joint actions and state information. When a global state exists, it directly uses it for training, otherwise, it uses joint action observation history for training. In this setting, the critical can get the information from the global. In the actor critical framework, the critical is often used to guide the learning of each agent and transmit the global information to each agent, so as to improve the modeling ability of each agent to the information of other agents (show in the Figure 3).

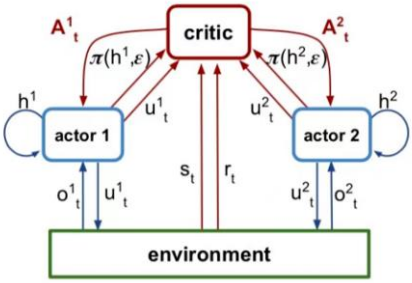

Figure 3. How COMA work

3.3. Resume game from death

Sometimes the agent will encounter some troublesome situation (show in the Figure 4), whatever agent chooses to click or do nothing will lead to crushing at the end. Although it's not common, one occurrence can be fatal. The solution to this can be when a bird is crushed, it can resume or go back to the last tube it just passed, in order to come up with a new way to pass this tube in this case after the resume.

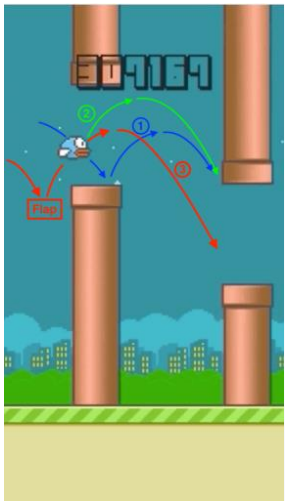

Figure 4. The troubles agents will face

4. Result and discussion

The result has shown that with the help of reinforcement learning, AI can be trained to play a game so well, even better than humans. From the original game, for people it is difficult to achieve 45 score (show in Figure 5), but after applying reinforcement study, the 1744 and even 10000 score can be easily achieved (show in Figure 6). At the beginning of our experiments, Flappy Bird's agent learning process was slow and required constant trial and error. This led to a slow growth in the first few scores. However, as the agent continues to trial and error and summarize, in the later stages the agent can already reach the score regularly and effectively gradually. Furthermore, all games elements are replaced including birds, obstacles (pipes), backgrounds, ground, and titles. And agent is redesigned, that little elf thing we'll call the Bluer, and Tubes are replaced by a chainsaw, which would make the game a better experience for the player, especially when combined with the apocalyptic setting (show in the Figure 7). Ultimately, the game presented was a brand-new game with the same gameplay as flappybird, called "flybluer". Although what have done at this stage is only the replacement of game elements, it is still a competition to the original game. This means that when people want to play a game that tests their reflexes, they face two choices, flappy bird and flybluer. This is how the market of similar games began to compete with each other, developing more interesting gameplay to attract more players, and eventually expanding the entire game market. Therefore, the flybluer can consider becoming a commercial game by adding more gameplay and features in the future. At the same time, AI are retrained to run through chainsaws with ease on our new game, and never die. Overall, our results show that deep reinforcement learning is a step in the right direction and has a lot of potential for further application

Figure 5. score before using reinforcement study

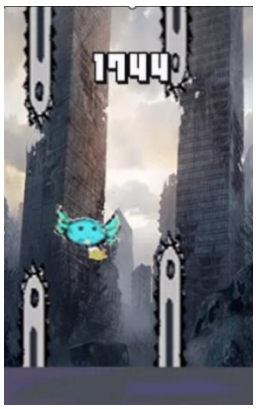

Figure 6. Score after using reinforcement study

Figure 7. The "flybluer"

5. Conclusions

The result has shown that with the help of reinforcement learning, AI can be trained to play a game so marvelous and fantastic, and even it performance is better than some professional gamers. For ordinary people, it is hard to achieve 100 score, but if the reinforcement study used in it, it may pretty be easy to achieve the target and even achieve higher score. In addition, although the layout, such as the agent, background, and the pipe, of the game is a bit outdated, after redesigning, it is full of fashionable elements, like blue agents, electric saw, and a doomsday atmosphere really attracted a lot of people to play, which creates a science fiction atmosphere to play. In addition, there is a probability to simplify the training process, but it still needs further experiments to prove it. Hopefully, in the future, the suggestion can be sufficiently offered to improve the project and make more improvements in training Artificial Intelligence (AI).

References

[1] Synopsys. What is reinforcement learning [R]. 2022 Https://www.synopsys.com/ai/what-is-reinforcement-learning.html

[2] Chen K. Deep reinforcement learning for flappy bird [R]. Stanford, 2015.

[3] Wikipedia. Flappy bird [R]. Https://en.m.wikipedia.org/wiki/Flappy_Bird, 2022.

[4] Naddaf Y. Game-independent ai agents for playing atari 2600 console games [R], 2010.

[5] Zhou Z H. Learnware: on the future of machine learning [J]. Frontiers Comput. Sci. 10.4 (2016): 589-590.

[6] Al-Jarrah O Y et al. Efficient machine learning for big data: A review [J]. Big Data Research 2.3 (2015): 87-93.

[7] Peteiro B D et al. A survey of methods for distributed machine learning [J]. Progress in Artificial Intelligence 2.1 (2013): 1-11.

[8] Qiu Y et al. Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training [J]. Biomedical Signal Processing and Control 72 (2022): 103323.

[9] Najafabadi M M, et al. Deep learning applications and challenges in big data analytics [J]. Journal of big data 2.1 (2015): 1-21.

[10] Zhang Q, et al. A survey on deep learning for big data [J]. Information Fusion 42 (2018): 146-157.