

Flappy Bird Game Based on Reinforcement Learning Q-Learning Algorithm

Zhenni He ^{1,†}, Yi Zhang ^{2,†}, Dingle Zhao ^{3,*}

¹ Wuxi United International School, Shanghai, China

² Beijing NO.80 Middle School, Beijing, China

³ RDFZ Chaoyang branch school, Beijing, China

* Corresponding Author Email: zhaodingyue@rdfzcygj.cn

[†]These authors contributed equally.

Abstract. In the field of Artificial Intelligence (AI), game AI is becoming more and more important, and the human-machine training of games is gradually driving the development of the game field. Among them, Flappy Bird is one of game which can controlled by an AI, which deserves more attention. In this work, we used Q-Learning as our main algorithm of the AI. In the flappy bird AI, the algorithm of Q-learning is used for giving the feedback through the environment which corresponding reward according to the actions of the agent. By using this method and after the training of the flappy bird AI, we can get the scores that are much more than human's record. The highest record of the flappy bird AI is 4, 083. The average score for human is about only 100, but in the flappy bird AI, the score can easily be more than 1, 000. According to all the work we did and all the result we got, we can see that the comparison between the AI and human. In the game area, AI did much better than human in most game. That is the reason that much research is focusing on developing game AI to help us getting deeper in the game field since it is more efficient to use.

Keywords: Reinforcement Learning, Flappy Bird, Machine Learning.

1. Introduction

The project the study is doing is that a Flappy Bird Clone using python-pygame. Flappy bird is a game that players need to manipulate a bird flying through the blocks made by water pipes which are in different length [1-3]. In the project, the technology of machine learning is needed to make sure the computer could play Flappy Bird automatically. The basic theory of the project is that by using different objects in the image to train convolutional neural network which is a widely used neural network [4-7], the scene state analysis based on the game screen and image recognition and classification are carried out. A pattern recognition procedure for particular states in the game scene is what it takes to determine the characteristics of the game from the original pixels and decide to conduct the appropriate actions. Furthermore, combing the artificial intelligence and the game is very important because it could practice the machine and the program, and since machine is good at analyzing possibilities better than human. For example, in the chess game, AlphaGo could win the human world champion easily by calculate thousands of possibilities that human could not accomplish in seconds [8-10]. We could use this technology in our life to help us predict different results that may appear.

A lot of the games that people are making today, they do not have any intelligence of their own, and people are starting to figure out how to make programs play games, and that involves machine learning, deep learning, or reinforcement learning. Many of the games we've heard about are controlled by AI in these ways. For example, snake AI implemented using A-star Pathfinding algorithm which intended to calculate the most optimal path to a target of the snake. 2048 AI implemented using Minimax algorithm and alpha-beta pruning, Minimax can be optimized by reducing the amount of processing and searching by alpha-beta pruning. In order to simulate the game between humans and AI, Minimax is a two-pass search where the first run assigns heuristic values to the nodes at the lowest depth and the second pass propagates the values up the tree. A Tetris AI

implemented using Pierre Dellacherie algorithm, it tries to perform each rotation of the current drop, placing it from left to right, to produce all the swing methods. Looking through all the evolution of various AI algorithms in various games, we wanted to complete Flappy Bird's self-training through reinforcement learning.

Our goal is for the bird to score as many points as possible, and to simplify the process, we can understand that the longer the bird lives, the better. To keep track of the predicted Q values for various actions in various states, the Q learning algorithm employs a Q table. The Q table will be initially initialized at random before the environment is explored. The agent will iteratively update Q (S, A) using the Bellman equation as it explores the environment. With the increase of the number of iterations, the agent will know more about the environment and the Q function can be fitted better and better. Until it converges or reaches a set number of iterations. The main advantage of Q learning is that it uses the time difference method TD (which integrates Monte Carlo and dynamic programming) to carry out offline learning, and Bellman equation can be used to solve the optimal strategy for Markov process.

2. Method

2.1. Introduction for Flappy bird

Flappy bird was a game developed by Dong Nguyen, and it was released on May 24, 2013. In the game Flappy Bird, players need to manipulate the bird by tipping the screen so that the bird could avoid touching the tubes. Additionally, Flappy Bird computer version is implemented by HTML5+Canvas and Javascript technology, which can be perfectly presented in FireFox and Chrome browsers. About its inventor, Dong Nguyen used to be a programmer who helped locate the taxis, and in weekends Nguyen would use his time producing a new game. He hoped that this game could be simple but challenging. As a result, he produced a game that players were asked to make the bird flaps its wings. Until January, 2014, Flappy Bird were downloaded fifty million times around the world.

2.2. Employed Q-Learning Algorithm

When we are completing our flappy bird AI, there are some methods we had used in this AI. The most significant one that was incorporated into the AI for Flappy Bird was Q-Learning. A reinforcement learning technique called Q-Learning will determine the subsequent optimal course of action given the current situation. It selects this action at random with the intention of maximizing the reward. Figure 1 presents the schematic diagram of the Q-Learning.

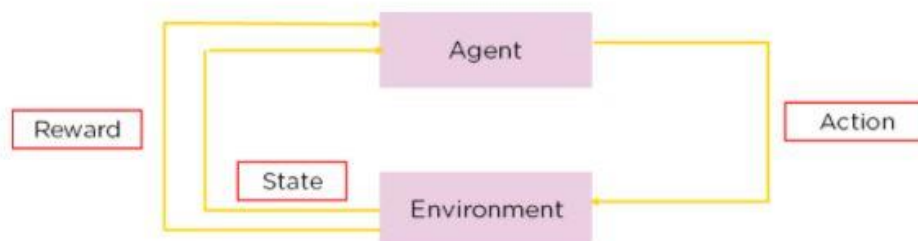


Figure 1. The schematic diagram of the Q-Learning.

It uses off-policy reinforcement learning without a model to determine the optimum course of action based on the agent's present condition. The agent will choose what to do next based on where it is in the environment. The model's objective is to determine the optimum course of action given the situation. It may do this by establishing its own regulations or by acting contrary to declared policies. This means that no policy is actually required, so we call it off-policy. Model-free refers to the use of forecasts of anticipated reactions to the environment by the agent to advance. Trial and error is used instead of a reward system for learning. Connecting to the flappy bird AI, we used Q-Learning as the

main algorithm. In our model, we used horizontal and vertical distances from the next pipe and vertical velocity of the bird to the state space to define the state of the bird. So that the bird can successfully get through the gap between the pipes. According to the flappy bird AI, the algorithm of Q-learning is used for giving the feedback through the environment which corresponding reward according to the actions of the agent. In order to store the Q value, the State and Action are built into a Q-table. Then, the Action that can yield the highest profit is chosen based on the Q value.

3. Results and discussion

In this study, the rule for the reinforcement learning is as follows: 1) Get one point for surviving 2) Death is awarded minus a thousand points. Every time we get a point, it's a screen. In other words, if we get 30 points for every second, we survive in a 30-frame game, then we get 300 points for every 10 seconds we survive, which is 1000 points for every death.

The state is based on how far each bird is from the next lower tube. We know that every time the bird or the tube moves, it doesn't have to be +1 every time, it has to be +4 or -4, so we can theoretically divide the number of states by 16 (x over 4, y over 4). The figure of the flappybird interface can be found in Figure 2.



Figure 2. The interface of the flappy bird game.

4. Conclusion

In this study, our aim is mainly to make a small game of Flappy Bird based on the reinforcement learning. Q-Learning was chosen in this study to make the bird fly better in the complete game. The experimental results demonstrated the feasibility of the employed method. In the future, we can also use the reinforcement learning algorithm to make more games and learn the code to facilitate some applications of human beings. However, we also have some shortcomings in this study, for example, the bird may fail at the beginning of the game and get a low score. Therefore, the modified and better version of the reinforcement learning algorithm can be considered.

References

- [1] Vu, T, FlapAI bird: training an agent to play flappy bird using reinforcement learning techniques [R]. arXiv preprint arXiv:2003.09579, 2020.
- [2] Ebeling-Rump, M, Applying q-learning to flappy bird. Department Of Mathematics And Statistics [D], Queen's University, 2016.
- [3] BasavaRaju, K, Flappy Bird Automation using TensorFlow [R]. Available at SSRN 3769788, 2021.
- [4] Yu, Q, Improved denoising autoencoder for maritime image denoising and semantic segmentation of USV [J]. China Communications, 17(3), 2020, 46-57.
- [5] Cai, G., Privacy-preserving CNN feature extraction and retrieval over medical images [J]. International Journal of Intelligent Systems, 2022.
- [6] Iqbal, S., A Heteromorphous Deep CNN Framework for Medical Image Segmentation Using Local Binary Pattern [J]. IEEE Access, 10, 63466-63480, 2022.
- [7] Jabloncik, F, SVM texture classification and R-CNN approach on medical image [J]. In 2022 ELEKTRO (ELEKTRO) (pp. 1-4). IEEE, 2022.
- [8] Carchidi, V. J. Do submarines swim? Methodological dualism and anthropomorphizing AlphaGo [J]. AI & SOCIETY, 1-13, 2022.
- [9] Pasqualini, L, Leela Zero Score: a Study of a Score-based AlphaGo Zero [R]. arXiv preprint arXiv:2201.13176, 2022.
- [10] Deeg, J. Schlaue, schlauer, am schlauesten: AlphaGo Zero [J]. In Künstliche Intelligenz (pp. 83-85). Springer, Berlin, Heidelberg, 2022.