Deep Reinforcement Learning in Atari Games

By

Andre Leonardo Angkawijaya

001201600010

A Thesis Proposal

Submitted to the Faculty of Computing

President University

in Partial Fulfilment of the Requirements

For the Degree of Bachelor of Science

in Information Technology

Cikarang, Bekasi, Indonesia

September, 2019

Deep Reinforcement Learning in Atari Games

By:

Andre Leonardo Angkawijaya
010201600010
Information System
President University

Approved by:

_____

Dr. Tjong Wan Sen, S. T., M. T.

Thesis Advisor

Abstract

This bachelor thesis's proposal deals with the implementation of *reinforcement learning algorithm* in a simulated environment, Atari games. Using games (simulated environment) as a development environment to improve *reinforcement learning algorithm* is a common practice to uses before the agent can actually be used on real life problem. Moreover, games offer a high-dimensional state data which is also a common problem to be solved in machine learning. The implementation of popular *reinforcement learning algorithm* called Deep Q-Learning will be implemented, tweaks to get optimum results, and benchmarked with the other existing *reinforcement algorithm*.

Problem Statement

Kuder, D., Hans, S., & Mittal, N. (2019) stated that AI will become a powerful business tools that could help people to win at business. In order to studies recent surge of AI trends in business, I intent to studies and compares different kinds of *reinforcement learning* algorithm which can train an *agent* to learn and interact within the specified environment to reach a specified goal optimally.

Introduction

Technology is developed by human to help in performing a complicated works that either involved dangerous works or complex computation. Through years, human proved that technology can also be utilized to assist human in their daily lives. Inventions such as computer and smartphone are some good examples of technologies development that enables human to work in a smart, simple, and efficient manner through a variety of smart programs. An example of the smart program is the virtual intelligence assistant developed by Google which can recognize our voice that can be processed as an input, the *Google Assistant.*

To prosper the quality of human's life, human began to developed a man-made intelligence, or what we usually called *Artificial Intelligence (AI)*. The long development goal of AI is to achieve the ability for the machine to *think and act* both *rationally and humanly* in solving any intellectual human task, *Artificial General Intelligence (AGI)*. However, in this study, we focused on building a something that perceives and acts, which will be called *agents*, that are able to *think and act rationally.* As AI's development still cannot reached the good performance of an AI that can *think and act humanly*.

Human ability to read a complex book is achieved by reading a simpler book, then, human began to gain knowledge and information to understand the complex one. Similar with human, machine receives inputs, calculates, and then show the predictions of the input. carving an intelligence into machine needs an *iteration of learning process* which is called *Machine Learning (ML).* In this section, the author describes four popular ML methods which called *Supervised, Unsupervised, Semi-supervised, and Reinforcement learning* according to Musumeci et al. (2018)

Supervised learning (SL) uses labelled/named data to trains the agents to predicts something, for example, the agent is trained with a labelled fruit images to be able to differentiates fruit's name when it receives a fruit image. In the *Unsupervised Learning (UL),* the agents are trained with an unlabelled data to find the pattern and classify the provided data. Market research, social network analysis, and data clustering are the example of the agents who used this training method. Agents trained using *Semi-supervised Learning* have the same purposes as SL, however it receives both labelled and unlabelled data.

The last training method, *Reinforcement Learning (RL),* will be the centre of this study. RL applies the *trial and error* learning method, where the agents learn the consequences of their

known *actions* in a specific environment (Andrew, N. G., n.d.). At the end of their *actions,* the *state* of the environment is evaluated. The agents are given a *reward* according to the *environment's state* which can be either *positive* or *negative,* the *positive* reward shown that the *agent's actions* satisfy our requirement whereas the *negative* reward do the opposites.

Google DeepMind and OpenAI are companies which utilizes RL in creating an expert agent that outperforms humans in game. Google DeepMind specializes in creating computer program which plays the game of Go, the *AlphaGo.* The program's successor, which is called *AlphaGo Zero,* have taught itself to play the game of Go for three days raining with only basic rules of Go as its base knowledge. It is reported that in Google DeepMind 2016 Challenge Match, the 18-time world champion Lee Sedol is defeated by *AlphaGo Zero* ("The Google DeepMind Challenge Match", n.d.). On the other hand, OpenAI Dota 2 bots also able to defeat three best Dota 2 player in the world in 1v1 match and it puts a tough battle in 5 bots vs 5 players mode ("OpenAI Five", n.d.).

The recent surge of AI in business become one factor which prove that AI's capability in predicting, clustering and classifying data, and organizing strategies is decent. Copeland (2016), a Silicon Valley Journalist writes for NVIDIA, the biggest graphic card companies, believes that it will not only be technology-driven businesses such as Google, Microsoft, and Amazon that utilizes AI. However, another business fields such as sports, oil, personal loans, and other companies will also utilizes AI to help them wins the business. Copeland cite Caulfield's (2015), NVIDIA's chief blogger, report of a beer's business which utilizes machine learning to helps craft brewers crafting a better beer by gaining a knowledge from their customers.

Hence, the author takes this chance to study *reinforcement learning* in his thesis to achieve bachelor degree entitled, *"The Application of Deep Reinforcement Learning in Atari games"*. The author will utilize the simulated environment, which is games, in learning RL algorithm and benchmark several (two or three) components to reach the agent's optimum performance in solving the Atari games.

Related Work

The most popular paper that the author finds throughout his research in this topic is the paper called "*Playing Atari with Deep Reinforcement Learning*" by Mnih et al. (2013). In this paper, the team introduces the implementation of Q-Learning variant into Deep RL algorithm, which are Deep Q-Network and Deep Q-Network Best, to create a single agent that is capable to achieve the highest score from seven different Atari Games. In their report, Mnih's team included their *reinforcement learning* algorithm which is shown in the *Figure 1* below.

Algorithm 1 Deep Q-learning with Experience Replay

Initialize replay memory $\mathcal{D}$ to capacity N
Initialize action-value function Q with random weights
for episode $= 1, M$ **do**
  Initialise sequence $s_1 = \{x_1\}$ and preprocessed sequenced $\phi_1 = \phi(s_1)$
  for $t = 1, T$ **do**
    With probability ϵ select a random action a_t
    otherwise select $a_t = \max_a Q^*(\phi(s_t), a; \theta)$
    Execute action a_t in emulator and observe reward r_t and image x_{t+1}
    Set $s_{t+1} = s_t, a_t, x_{t+1}$ and preprocess $\phi_{t+1} = \phi(s_{t+1})$
    Store transition $(\phi_t, a_t, r_t, \phi_{t+1})$ in $\mathcal{D}$
    Sample random minibatch of transitions $(\phi_j, a_j, r_j, \phi_{j+1})$ from $\mathcal{D}$
    Set $y_j = \begin{cases} r_j & \text{for terminal } \phi_{j+1} \\ r_j + \gamma \max_{a'} Q(\phi_{j+1}, a'; \theta) & \text{for non-terminal } \phi_{j+1} \end{cases}$
    Perform a gradient descent step on $(y_j - Q(\phi_j, a_j; \theta))^2$ according to equation 3
  end for
end for

Figure 1. The algorithm for Deep Q-Learning with experience replay by Mnih et al. (2013)

The Deep Q-Learning with experience replay algorithm will be the main algorithm that the author will try to implements and tweaks in his work. Additionally, the other related works that the author found is the bachelor thesis entitled *Deep Q-Learning with Feature Exemplified by Pacman* by Meo (n. d.). In his thesis, Meo studies and implement Deep Q-Learning in PacMan game. Furthermore, Meo experiments between different training setup and algorithm to benchmark and retrieve the most satisfying performance of his algorithm.

Mnih and Meo's works are the base of the author's thesis. In his work, the author would like to take a single main algorithm and then compares it with different type of existing *reinforcement learning* algorithm. Then, the author would tweak different scenario in his setting and record it to search for the satisfying performance that his algorithm implementation can deliver.

Problem Description

The uses of games environment as the means to simulated AI learning progress is a popular practice in training AI, especially to trains the AI that will take on a risky task. For example, the intelligence mechanical hand robot which have task on moving heavy or fragile object. If the training were to be conducted in real life, the business will only suffer a lot of damage through the training process. A lot of product or placeholder will be wasted just to train the hand's capabilities and the hand itself could be damaged through a lot of iterations of learning process.

Thus, game is utilized as it simulates real-life environment which can be used to trains AI to minimalize costs and risks in order to achieve greater task. The author utilizes M. G. Bellemare, Y. Naddaf, J. Veness, & M. Bowling (2013) creation of *Arcade Learning Environment (ALE)* which is a dedicated simple object-oriented framework for hobbyists and AI researchers to developed AI agents using Atari games as shown by the *Figure 2* below. Additionally, the author uses OpenAI Gym, which is an enhanced toolkit for creating an agent trained by RL which uses ALE.

Figure 2. Gym Retro screenshots collage showing Atari and Sega games environment. Reprinted from Gym Retro, in *OpenAI*, 2018, Retrieved from https://openai.com/blog/gym-retro/. Copyright 2018 by OpenAI. Reprinted with permission.

By utilizing OpenAI, the author will gain an access to a lot of documentation that could help the author in building his first agent. A lot of papers are also available that could give the author an insight for his job.

Solution Strategy

OpenAI research in RL through various papers line up a nearly accurate taxonomy of algorithms in modern RL as shown by *Figure 3* below. In this study, the author which uses will utilizes a Model-Free RL algorithm, specifically the Deep Q-Networks (DQN) and/or Categorical 51-Atom DQN (C51), a variant of DQN.

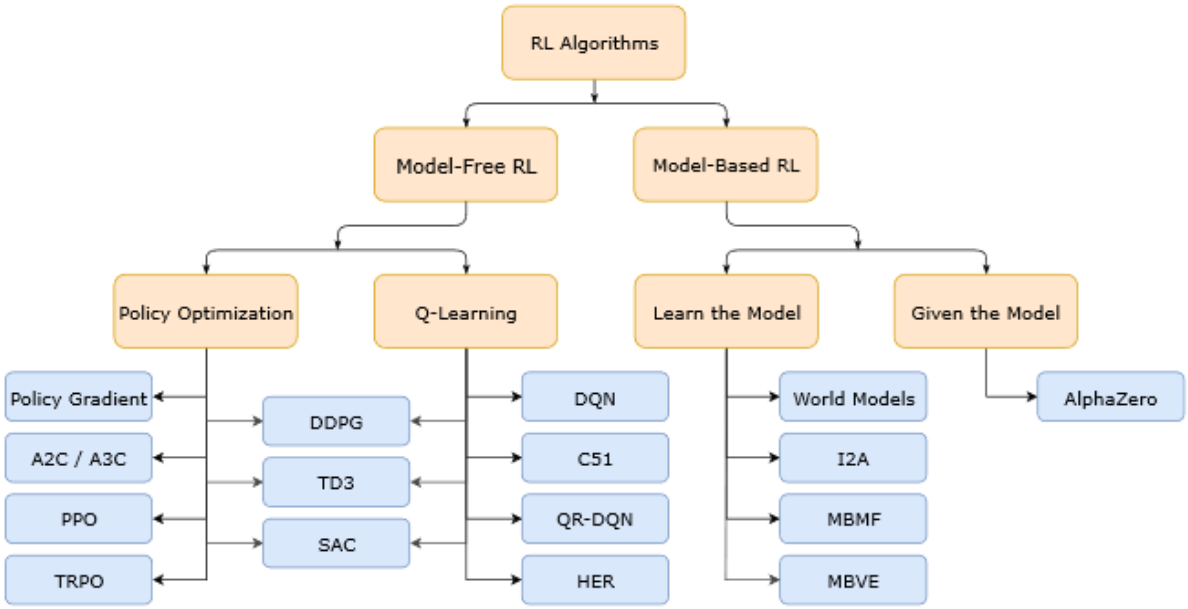

Figure 3. A non-exhaustive, but useful taxonomy of algorithms in modern RL. Reprinted from Part 2: Kinds of RL Algorithm, in *OpenAI Spinning Up*, 2018, Retrieved September 12, 2019, from https://spinningup.openai.com/en/latest/spinningup/rl_intro2.html. Copyright 2018 by OpenAI. Reprinted with permission.

A Model-Free RL specified that the agents do not have full observation to the environments. In one screen alone, the agent can only observe a partial information of Atari games environment (Hausknecht, M., & Stone, P, 2018). For example, in the screen of the game of the game of Pong, the game only reveals two paddles and the ball, however the velocity for the ball is unknown to the agents as shown by the *Figure 4* below. Therefore, the author does not need to fully feeds the environment's model within the game.

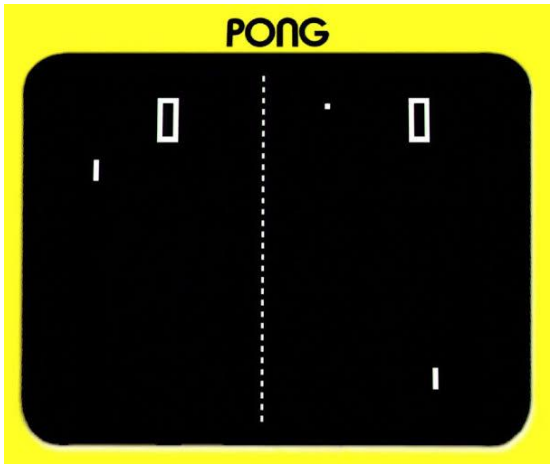

Figure 4. Pong. Reprinted from Pong, in *Silverball Museum*, n.d., Retrieved from http://silverballmuseum.com/product/pong/. Copyright 2012-2016 by Avada. Reprinted with permission.

In his project, the author will firstly research his approach to develops the model by using the selected algorithm mentioned in the paragraph above (might be different when the author found a better approaches). Next, the author will prepare the data pre-processing stages (normalizing the needed data for the training). Then, the author will try to build the algorithm and begun the agent's training for around a thousand up to ten thousand iterations. It should be noted that the author will implements the algorithm manually (the author will not use the built-in algorithm provided by the framework as it will destroy the purpose of thesis).

Using this widely and common uses model will help the author in learning and building the agent trained by RL in the expected duration of finishing the bachelor's thesis as expected by the committees. As there exist a lot of papers and documentations that can give a great help for the author to build his own agent. The author also will get an additional new knowledge in AI's fields, as the author have already grasped the concept of AI and experienced the basic of SL methods through the classes in the previous semesters.

Evaluation

Through the works the author has described in the previous section, the author will take the evaluation progress through benchmarking. As mentioned, OpenAI Gym is a framework that already gives developer the built-in function to call RL's algorithm. In this project, OpenAI Gym built-in function will become one evaluation tool for the author's RL algorithm by the mean of comparison.

Implementing different kinds of algorithm to benchmark the author's algorithm will take a lot of time. Thus, to fulfil the tight deadline of bachelor's degree thesis submission, the author will use the benchmarking of another learning method taken from the other research papers. For example, the popular "*Playing Atari with Deep Reinforcement Learning*" evaluation by Mnih et al. (2013). Their evaluation, which can be seen in *Table 1* below, consist of them benchmarking seven different Atari games with various kind of algorithm where DQN is their main algorithm.

Table 1.

The algorithm benchmarking of "Playing Atari with Deep Reinforcement Learning" research

	B. Rider	Breakout	Enduro	Pong	Q*bert	Seaquest	S. Invaders
Random	354	1.2	0	−20.4	157	110	179
Sarsa [3]	996	5.2	129	−19	614	665	271
Contingency [4]	1743	6	159	−17	960	723	268
DQN	**4092**	**168**	**470**	**20**	**1952**	**1705**	**581**
Human	7456	31	368	−3	18900	28010	3690
HNeat Best [8]	3616	52	106	19	1800	920	**1720**
HNeat Pixel [8]	1332	4	91	−16	1325	800	1145
DQN Best	**5184**	**225**	**661**	**21**	**4500**	**1740**	1075

Note. DQN = Deep Q-Network, From *Playing Atari with Deep Reinforcement Learning*, by Mnih et al. (2013).

Another step of benchmarking that the author can also take is to use the available built-in function from OpenAI Gym to run another RL algorithm. Utilizing this built-in function will be powerful tools for the author to compare his work results.

Required Resources

The following points specified the required resources to develop the thesis's project:

1. Laptop or computer with Intel Core i7 and high-end NVIDIA Graphic Cards (above or equals GTX 950M)
2. PyCharm as the Integrated Development Environment
3. Anaconda, a Python data science platform program
4. OpenAi Gym, a toolkit for developing reinforcement learning algorithm
5. Atari-Py, a Python binding to Atari games

Thesis Timeline

No.	Activity Name	Duration	Start Date	End Date
1.	System Planning	1 week	September 3, 2019	September 10, 2019
2.	System Analysis	1 week	September 10, 2019	September 17, 2019
3.	System Design	3 weeks	September 17, 2019	October 8, 2019
4.	System Implementation	7 weeks	October 8, 2019	December 4, 2019
5.	System Testing	2 weeks	December 4, 2019	December 18, 2019

Summary

The authors find out through his research that the recent achievement of AI agents that were trained by *reinforcement learning* methods shows a promising advancement of AI. The author believes that researching AI shows a lot of promises in improving a lot of sector in human life. Showing the interest in learning RL algorithms, the author utilizes arcade game framework and determines to conduct deep research into *deep reinforcement learning algorithm* and tweaks several parts of the algorithm to get the most optimum result.

Bibliography

Andrew, N. G. (n.d.). Part XIII Reinforcement Learning and Control [PDF document]. Retrieved from CS229 Stanford Edu Website: http://cs229.stanford.edu/notes/cs229-notes12.pdf

Achiam, J., & Morales, M. (2018). *A non-exhaustive, but useful taxonomy of algorithms in modern RL*. [Graph]. Retrieved September 12, 2019, from https://spinningup.openai.com/en/latest/spinningup/rl_intro2.html

Caulfield, B. (2015, September 2). Better Beer Through GPUs: How GPUs and Deep Learning Help Brewers Improve Their Suds [blog post]. Retrieved from https://blogs.nvidia.com/blog/2015/09/02/beer/

Copeland, M. (2016, October 17). From Winning Go to Making Dough: What Can Deep Learning Do for Your Business? [blog post]. Retrieved from https://blogs.nvidia.com/blog/2016/10/17/deep-learning-help-business/

Hausknecht, M., & Stone, P. (2015). Deep Recurrent Q-Learning for Partially Observable MDPs [PDF document]. In *AAAI Fall Symposium on Sequential Decision Making for Intelligent Agents (AAAI-SDMIA15)*, Arlington, Virginia, USA, November 2015. Retrieved from https://www.cs.utexas.edu/~pstone/Papers/bib2html-links/SDMIA15-Hausknecht.pdf

Kuder, D., Hans, S., & Mittal, N. (2019). AI-fueled organizations Reaching AI's full potential in the enterprise. Retrieved from https://www2.deloitte.com/us/en/insights/focus/tech-trends/2019/driving-ai-potential-organizations.html

M. G. Bellemare, Y. Naddaf, J. Veness, & M. Bowling. *The Arcade Learning Environment: An Evaluation Platform for General Agents, Journal of Artificial Intelligence Research, Volume 47, pages 253-279, 2013.*

Meo, R. (n. d.). *Deep Q-Learning with Feature Exemplified by Pacman* (Bachelor's Thesis, Hamburg University of Applied Sciences). Retrieved from http://edoc.sub.uni-hamburg.de/haw/volltexte/2018/4168/pdf/Roland_Meo_BA_Thesis.pdf

Mnih et al. (2013, December 19). *Playing Atari with Deep Reinforcement Learning* [PDF]. Retrieved from https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf

Musumeci et al. (2018). An Overview on Application of Machine Learning Techniques in Optical Networks [PDF document]. Retrieved from https://arxiv.org/pdf/1803.07976.pdf

OpenAI Five. (n.d.). Retrieved from https://openai.com/five/

Pfau et al. (2018). *Gym Retro screenshots collage showing Atari and Sega games environment*. [Screenshots collage]. Retrieved from https://openai.com/blog/gym-retro/

The Google DeepMind Challenge Match. (n.d.). Retrieved from https://deepmind.com/alphago-korea