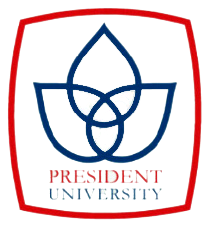
THESIS PROPOSAL



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:

Dilla Putri Aprilia

012201600007

Information System

President University

Approved by:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

Thesis Advisor

**Abstract**

**Problem** **Statement**

*// Mentioning AGI might be redundant, instead, mention AI impacts to business*

Reaching the state of *Artificial General Intelligence (AGI)* which enables *Artificial Intelligence (AI)* to solve general intellectual human task is a long development goals of AI. In order to studies the recent AI trends in business that could possibly be the path to reach AGI, I intent to use *reinforcement learning* which can train an *agent* to learn and interact with the specified environment to reach a specified goal optimally.

**Introduction**

*// Paragraph one might be redundant*

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\**

*// Should create better introduction to AI if paragraph one is deleted*

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, which called *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*.

*// insert pictures and references on AGI*

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)\**. There exist four ML methods which called *Supervised, Unsupervised, Semi-supervised, and Reinforcement learning*.

*// insert ref for ML description.*

*// The following sections describing other ML methods other than ML might be redundant*

*Supervised learning* uses labelled/named data to trains the agents to predicts something, for example, the agents 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,* 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 *supervised learning*, however it receives both labelled and unlabelled data.

*// insert pictures and references on both SL and UL*

The last training method, *Reinforcement learning,* will be the centre of this study. *Reinforcement learning* applies the *trial and error* learning method, where the agents learns the consequences of their known *actions* in a specific environment. 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*. ///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.

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. 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, but soon another business fields such as sports, oil, personal loans, and other companies will also utilizes AI to help them wins the business.

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 *reinforcement learnings* algorithm and benchmark several (two or three) components to reach the agent’s optimum performance in solving the Atari games.

**Related Work**

**Problem Description – PROB DETAILS on ENV, EXMPLE, WHY IS IT MATTERS? FROM RL Prob -> SIMULATION**

*// The paragraph is still not powerful enough to persuades the committees*

*// TODO: Clearly describes the real-world problem where AI impact business*

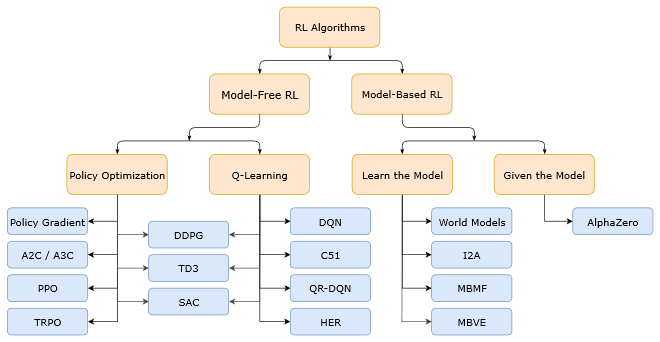
An AI mechanical hand robot which have task on moving heavy object automatically powered by AI will suffer a lot of damage and loss if the training is conducted in a real world. Thus, game is utilized as it simulates real-life environment which can be used to trains AI to minimalize costs and risks in order to achieve greater task. The author utilizes *Arcade Learning Environment (ALE)* which is a dedicated simple object-oriented framework for hobbyists and AI researchers to developed AI agents using Atari games.



*Figure x.* 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.

**Solution Strategy**

OpenAI research in RL through various papers line up a nearly accurate taxonomy of algorithms in modern RL as shown by *Figure x* below. In this study, the author which uses the ALE 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 x.* The nearly accurate non-exhaustive taxonomy of RL Algorithm. 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 access to the model of the environments. Atari simulator in ALE hid the source code implementation, which means the agents does not have the full information of the environment (partially observable environment). Additionally, the game does not give the agents a detail of the environment. In the screen of the game of Pong, the game only reveals two paddles and the ball, however the velocity for the ball is unknown. Which

**Evaluation**

The evaluation is conducted by

**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
6. Alt. Project malmo, a toolkit for developing reinforcement learning algorithm in Minecraft

# 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**

A brief (one to two paragraph) summary of the proposal (i.e. the previous sections) that highlights the key points in the proposal and provides a list of contributions to the field that you expect your work to provide. Be very specific when listing your contributions and explain why they are of interest to the computing community. .

**Bibliography**

Andrew, N. G. 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>

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.*

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>

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

Sutton, R. S., & Barto, A. G. (2014, 2015). *Reinforcement Learning: An Introduction* (2nd ed in progress). Retrieved from <https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf>