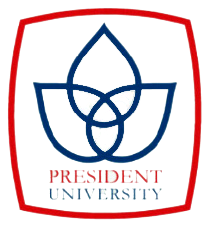
THESIS PROPOSAL



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:

Andre Leonardo Angkawijaya

010201600010

Information System

President University

Approved by:

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

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

Thesis Advisor

**Abstract**

**Problem** **Statement**

Reaching the state of *Artificial General Intelligence (AGI)* which enables *Artificial Intelligence (AI)* to solve general intellectual human task (daily task, business, academic, etc.) is a researcher’s long development goals. In order to studies recent surge of AI trends in business, I intent to use *reinforcement learning* which can train an *agent* to learn and interact within the specified environment to reach a specified goal optimally.

**Introduction**

*// Paragraph one might be redundant, might seriously consider deletes paragraph 1.*

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*. As AI’s development still cannot reached the good performance of an AI that can *think and act humanly*.

*// 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)*. 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)

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

**Related Work**

A related work section that surveys previous work related to what you are proposing. This section should be carefully written and organized to make the relationships between the earlier research efforts clear and to also explain how that research relates to your proposed work. It is primarily this section that makes it apparent to the committee that you are, in fact, prepared to undertake your proposed work. The work you reference should be quite extensive, relevant and recent. Insufficient references suggest to the committee that you may not be aware of all the related work and this means that it is possible that your work may already have been done by someone else. The inclusion of irrelevant (or too many) references may lead committee members to question your understanding of the area. Finally, lack of recent references might suggest that your proposed work is no longer of interest or is, perhaps, too hard a problem that other researchers have chosen to overlook. Finally, be careful to base your related work on quality publications. All (or very close to all) of your referenced papers should be from well-respected, refereed sources (i.e. journals or top tier conferences in your selected area). Referring to dubious papers lessens the committee’s confidence in your thesis proposal. Finally, your selected papers should reflect a reasonable amount of breadth in terms of authorship and source. Insufficient breadth might lead the committee to fear that you are following individual opinion instead of well-founded and widely accepted scientific results.

**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 x* below. Additionally, the author uses OpenAi Gym, which is an enhanced toolkit for creating an agents trained by *reinforcement learning* which uses ALE.

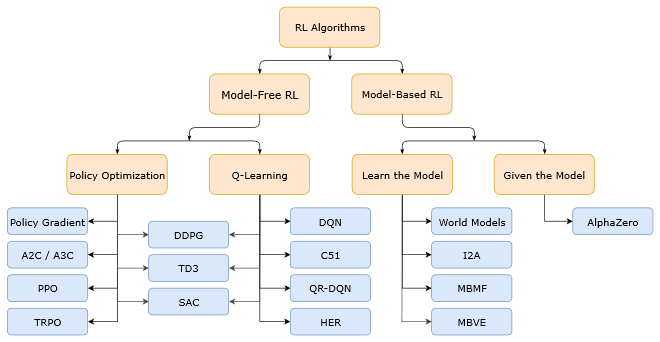


*Figure x.* Gym Retro screenshots collage showing Atari and Sega games environment. Reprinted from Gym Retro, in *OpenAI*, 2018, Retrieved from <https://openai.com/blog/gym-retro/>. Copyright 2018 by OpenAI. Reprinted with permission.

By utilizing OpenAI Gym, the author will gain an access to a lot of documentation that could helps 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 x* 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 x.* 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. Atari games does not contain a full observable information from one screen (Hausknecht, M., & Stone, P, 2018). For example, 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 to the agents as shown by the *Figure x* below. Therefore, the author does not need to fully feeds the environment’s model within the game.

In his project, the author will firstly research his approach to develops the model by using the selected algorithm mentioned 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 models and begun the agent’s training for around a thousand up to ten thousand iterations.

Using this widely and common uses model will help the author in learning and building the agent trained by *reinforcement learning* 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 *Supervised Learning* methods through the classes in the previous semesters.

**Evaluation -> related works comparison, by utilizing different activation function/optimizer/etc., tws guidance, each training iterations result will be recorded in a graph (show the example of other people papers),**

A description of how you propose to undertake the evaluation of your work. You must ultimately be able to answer the question of whether or not the work you have proposed and (later) completed is important. This is often done by direct comparison with other, existing work in the field. Such comparisons may be done experimentally, analytically, through simulation or possibly a combination of these. For example, you might be proposing a thesis where, at the end, you will want to compare the performance of an algorithm you developed and implemented with the performance of a similar existing algorithm. When doing this, always try to make the comparison(s) as objective and meaningful as possible. Compare your type of apple to someone else’s type of apple, not to an orange. Be sure to explain the methodology behind your comparison (e.g. how you will gather performance results accurately or how you will construct and run a simulation study). Always remember to keep **statistical** **significance** in mind whenever this is appropriate. Results based **on samples of small size do not constitute evidence of improvement nor do results where the degree of improvement exceeds the margin for error in the exper**iment. When actually doing the assessment, try to be totally objective and always **resist the temptation to tweak your work until you get the “expected” results.** Instead, explain the results you get.

**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 *reinforcement learning*, 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 research.

**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.*

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>

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>

<https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf> 2013 Volodymyr Mnih Koray Kavukcuoglu David Silver Alex Graves Ioannis Antonoglou Daan Wierstra Martin Riedmiller Playing Atari with Deep Reinforcement Learning