

NUS-ISS – Intelligent Software Agents  
Self-learning Systems Project

Intelligent Game-playing Agent Using  
Advanced Reinforcement Learning Algorithms

Koh Sook Bing (A0195413E)  
Lim Chong Seng Hermann (A0195392U)  
Yeo Whye Chung Nelson (A0195405A)

Table of Contents

[Table of Contents 2](#_Toc40676513)

[Executive Summary 3](#_Toc40676514)

[Problem Overview 4](#_Toc40676515)

[Technical Approach 5](#_Toc40676516)

[Results and Evaluation 6](#_Toc40676517)

[Conclusion 7](#_Toc40676518)

[References 8](#_Toc40676519)

# Executive Summary

ABCDEFG

# Problem Overview

## Problem Statement

ABC

## Project Objectives

In this project, the team will study the gameplay performance of three newer reinforcement learning algorithms developed after the Deep Q Network, one of the most popular algorithms today. The algorithms are:

* Synchronous Advantage Actor Critic (A2C)
* Hindsight Experience Relay (HER)
* Proximal Policy Optimisation (PPO)

This report will give a brief overview of each algorithm, discuss the observations made on trends and gameplay performance at different hyper-parameter configurations, and the comparative performance of optimised models for each algorithm.

## Game Overview

The game used in this study is named Space Invaders. First released as an arcade game during the 1970s, it is one of the earliest shooting games developed, and commonly considered to be the highest-grossing video game of all time. The gameplay mechanics is described as follows:

* The player controls a star-ship at the bottom of the screen, moving sideways or shooting upwards
* Five rows of enemy aliens start from the top, moving sideways back-and-forth and gradually advancing downwards, all while firing shots down towards the player
* The objective of the game is to shoot down all of the aliens, before they reach the player, scoring as high as possible, while avoiding incoming shots to survive
* When the ship is hit, the player will lose one of three lives. The game is over when all three lives are lost
* Each alien is worth a certain amount of points. The higher up in the back of the group an alien is, the more points it is worth.
* As more aliens are shot down, the remaining aliens gradually increase in speed, becoming progressively harder to hit
* Once all the aliens are shot down, the player progresses to the next level with a new set of aliens
* Occasionally, a special alien ship will fly sideways across the top of the screen. The player will earn a large amount of bonus points if the ship is shot down
* The bottom of the screen, above the player, is also comprised of four “bunkers” which can shield the player from incoming shots, although the enemy shots will gradually disintegrate the shelter.

# Technical Approach

## Overview

In this section, the report gives a brief description of the developed system’s technical implementation, and an overview of the three algorithms being studied.

## System Implementation

The system is implemented using two main Python libraries:

* gym-retro: Developed by OpenAI, this library supports emulation of Atari, NEC, Nintendo, and Sega game consoles, enabling creation of game environments for about 1,000 games, a much wider range than the classic gym library, which can only emulate Atari2600 games
* stable-baselines: This library is a fork from OpenAI’s baselines library, a module providing reusable implementations of several reinforcement learning algorithms for the purpose of simplifying the agent development process, so that developers can focus on environmental design and hyper-parameter optimisation. stable-baselines improves upon baselines with some major code refactoring and clean-ups, improved test coverage of the module’s codes, provision of better documentation, and implementation of several additional algorithms

## Synchronous Advantage Actor Critic (A2C)

Overview of algo

## Hindsight Experience Relay (HER)

Overview of algo

## Proximal Policy Optimisation (PPO)

Overview of algo

# Results and Evaluation

## Overview

In this section, the report looks at the hyper-parameter tuning process for each algorithm, discussing changes in gameplay performance and trends observed as hyper-parameter values were being tuned. Finally, models of the three optimised algorithms will be evaluated to compare their gameplay performance.

## Synchronous Advantage Actor Critic (A2C)

Overview of algo

## Hindsight Experience Relay (HER)

Overview of algo

## Proximal Policy Optimisation (PPO)

Overview of algo

## Final Evaluation

# Conclusion

ABC

# Bibliography

Dump all the links here