International Journal of Future Generation Communication and Networking
Vol. 13, No. 3s, (2020), pp. 1604–1607

Reinforcement learning in OpenAI five

Harsh Agrawal, Shweta B. Guja

Computer Science NBNSSOE

Abstract

OpenAI Five is a machine learning project developed by OpenAI that acts as a team of game bots to play against humans in the competitive computer game Dota 2. In the game Dota 2, "The International" is the world championship tournament. A game between Dota 2 world champions and game bots developed by OpenAI was organised and the result was mind blowing. The International world champion 'OG' lost back to back games against OpenAI Five. OpenAI five observes the game after extracting the present game state from Dota developer's API with one layer which contains 1024- unit LSTM. Without any human data involved, the neural network conducts actions via numerous available action heads and each head has meaning. In other words, bots play against each other learning best actions countering every possible move, in short opponent modelling. Opponent modeling is the strategy to observe and anticipate the moves and ability of an opponent. In gaming, the model is an abstracted description of the opponent and his strategy, countering the opponent's behavior in the game. This method is important in multiple agent settings where secondary agents with competing goals also adapt with their abilities [2], yet it remains challenging because strategies interact with each other and change. The result showed that AI (if trained in expert supervision) can be proved effective against a human mind in difficult situations.

Keywords *- artificial intelligence, OpenAI, Dota 2, opponent modeling, Proximal Policy Optimization.*

1. Introduction

Games are the best medium to create, develop artificial intelligence and compare it with the existing AI in order to improve the current computer intelligence power. Chess, checkers, and Go are examples of games frequently used for this task [1]. In the past many such artificial intelligences have shown extra ordinary behaviour by defeating World Champions in their own game. For example Google's AlphaGo AI, DeepBlue by IBM, OpenAI Five and many more.

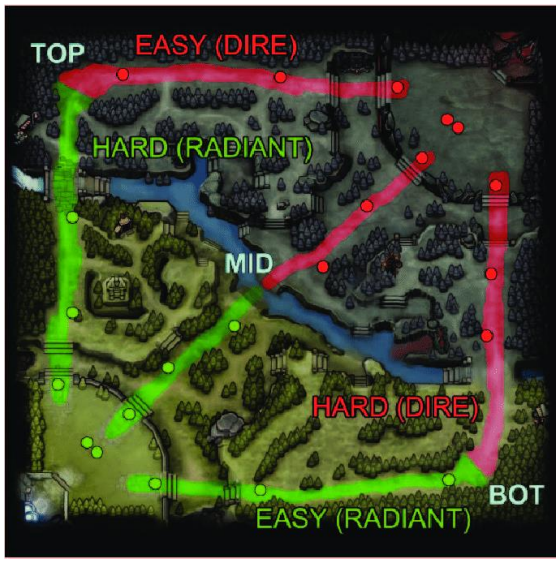
AlphaGo is the first computer program to defeat a human Go world champion, the strongest Go player in history. Deep Blue was a chess-playing computer made by IBM. It is known for being the first computer chess-playing system to win a chess game against a reigning world champion under regular time controls[6]. Recently there was a game of Dota 2 played by OpenAI bots and World Champions in which Who could have thought that artificial intelligence can defeat a team of five human minds in their own game which they played all their life.

Normal bots have been around since the game launch but their skills are limited. There are levels of bots, level 1 bot being the weakest and level 2 with increasing difficulty but in the bigger picture their skills have a limit, they can't exceed a particular threshold[7]. These special bots created by extraordinary companies have proven that this extent of levels can be broken and a new level of intelligence can be achieved.

ISSN: 2233-7857 IJFGCN

2. Dota 2 Architecture

Dota 2 is a multiplayer online battle arena (MOBA) video game made and released by Valve [1]. Dota 2 is game played between ten players with five players in one team and five on another, with each team protecting their own separate base on the map [2]. Every player in the game controls a strong character, known as a "hero", all heroes have unique abilities and different styles of play. In this PvP combat game a player collects experience points for their heroes in order to defeat opposing team's heroes. A team wins by being the first to destroy the other team's "Ancient", a large structure located within their base [3].



To properly play the game you need at least a year of learning time to actually know all the basic strategies of all the items in the game. To think that AI can master such a hard game in way less time than humans, strongly suggest that competing against robots can improve one's skills.

3. OpenAI Five Bots

Bots play 180 year worth of games against itself every day to learn and improve its abilities, it trains using a scaled-up version of Proximal Policy Optimization [3] using 256 GPUs and 128,000 CPU cores. It uses 1 LTSM for each hero with no human data and learns best strategies. This shows that reinforcement learning [5] can yield optimum output, contrary to our own expectations upon starting the project.

3.1 Progress

In this complex video game, AI's objective is to exceed human capabilities. Previously AI has successfully mastered simple games like chess and go but to conquer a game like dota it requires a lot of machine power, thanks to next gen GPUs and CPUs now it's possible. Observing every fourth frame OpenAI Five yields 20,000 moves for the hero in dota 2. Go usually ends before 150 moves, Chess before 40 moves, with almost every move strategic.

3.2 Problem

Dota games run for 45 min (average) resulting in 80,000 ticks per game. Most actions like moving a hero has minimal impact but some actions like using a town portal (used for teleportation) can affect the game

ISSN: 2233-7857 IJFGCN

strategically [1]. The map discretize the space into 170,000 possible actions per hero with an average of 1000 actions per tick [2]. In chess average number of actions is 35 and 250 in Go.

3.3 Solution

OpenAI system learns using a massively-scaled version of Proximal Policy Optimization. Without any data from human replays they start with random parameters.

CPUs	128,000 pre-emptible CPU cores on GCP
GPUs	256 P100 GPUs on GCP
Experience collected	~900 years per day
Size of observation	~36.8 kB
Observations per second of gameplay	7.5
Batch size	1,048,576 observations
Batches per minute	~60

3.4. How is it better?

These bots are trained using Opponent modeling strategy. It is a strategy (in reinforcement learning) to compete against another AI in a simulation battle until it reaches all possible outcomes. In opponent modeling first AI is taught the basics of the game then AI fights with itself in a continuous simulation battles until maximum possible strategies have been learned. This ensures that in difficult situations the best actions could be taken resulting in good outcomes. It is the high-level goal of powering the "best" or "least exploitable" decision for the opponent that is consistent with views of the opponent's behavior [4].By using this model OpenAI Five proved that Artificial Intelligence can defeat the greatest of minds the world has seen.

4. Overview

As the technology is advancing day by day, new machine learning techniques are invented. Dota 2 one of most popular and complex game with top quality AI requirement best presented by Font Fernandez in his paper “The Dota 2 Bot Competition” [1] simply gives the brief idea of AI in games.

Next major step was to accurately fuse intelligent agent (AI) with games using some sort of technique for results.

ISSN: 2233-7857 IJFGCN

Imtiyaz in his paper "*Implementation of Intelligent Agent in Defense of the Ancient 2 through Utilization of Opponent Modeling"* [2] showed opponent modeling, a reinforcement learning technique, to successfully combine AI with games.

Proximal policy optimization (reinforcement learning) a method used to train bots, OpenAI used this strategy to train all their bots and was able to defeat world champions in dota 2 game. A training technique published by Zhang, Z in "*Proximal Policy Optimization with Mixed Distributed Training" [3]* is easy to implement and most effective one out there.

5. Conclusions and Future work

Here I have presented that a robot when trained using Proximal Policy Optimization can master any provided field. In the previous example OG (World champions) lost against bots, if these bots are trained in other fields like sports or medicine then we can reach whole new level of advancement in technology. For example if a robot is taught how to play badminton and in simulation it is allowed to play itself then for sure it will learn amazing new strategies which will definitely enhance the skills of pros. A strategy which can outsmart the professionals in their fields, will shake the world from its core.

Refrences

[1] Font Fernandez, J. M., & Mahlmann, T. (2018). The Dota 2 Bot Competition. IEEE Transactions on Games, 1–1. doi:10.1109/tg.2018.2834566.

[2] Imtiyaz, A. H., & Ulfa Maulidevi, N. (2018). Implementation of Intelligent Agent in Defense of the Ancient 2 through Utilization of Opponent Modeling. 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA). doi:10.1109/icaicta.2018.8541286

[3] Zhang, Z., Luo, X., Liu, T., Xie, S., Wang, J., Wang, W., … Peng, Y. (2019). Proximal Policy Optimization with Mixed Distributed Training. 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI). doi:10.1109/ictai.2019.00206

[4] Farouk, G. M., Moawad, I. F., & Aref, M. M. (2017). A machine learning based system for mostly automating opponent modeling in real-time strategy games. 2017 12th International Conference on Computer Engineering and Systems (ICCES). doi:10.1109/icces.2017.8275329

[5] Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep Reinforcement Learning: A Brief Survey. IEEE Signal Processing Magazine, 34(6), 26–38. doi:10.1109/msp.2017.2743240

[6]Dhumane, A., & Prasad, R. (2015). Routing challenges in internet of things. CSI Communications.

[7] Dhumane, A. V., Prasad, R. S., & Prasad, J. R. (2017). An optimal routing algorithm for internet of things enabling technologies. International Journal of Rough Sets and Data Analysis, 4(3), 1–16.

ISSN: 2233-7857 IJFGCN