修 士 論 文 概 要 書

Summary of Master's Thesis

Date of submission: 01/29/2020 (MM/DD/YYYY)

専攻名 (専門分野) Department	Computer Science and Communications Engineering	氏 名 Name	Bundik Bettina Vivien	指 導 教 員 Advisor	菅原俊治 印 Seal
研究指導名 Research guidance	Research on Intelligent Software	学籍番号 Student ID number	CD 5118FG13-8		
研究題目 Title	Features and Performance of Sarsa Reinforcement Learning Algorithm with Eligibility Traces and Local Environment Analysis for Bots in First Person Shooter Games				

1. Introduction

Reinforcement learning agents embedded in first person shooter video games make a good test bed for applying artificial intelligence to complex environments. FPS bots share plenty of common features with robotics and multi-agent systems so this field of research is truly promising. In such an environment, basic skills need to be developed for navigating through a map of the game, collecting valuable items, engaging in combat with opponents and the ability to survive. General goals of FPS learning agents include minimizing collisions with geometry of the environment, maximizing distance travelled and number of items collected, increasing the count of kills and decreasing the count of deaths. In addition, if a variety of adaptive and effective behaviours of agents is achieved, bots will have unexpectedness and the overall game-play becomes more enjoyable. The purpose of this paper is to reach these goals and outperform previous FPS bot learning methods by a local environment analysis of agents.

2. Related Work

Previous studies prove that applying tabular Sarsa(λ) algorithm with eligibility traces to agents in an FPS environment is effective and well-working. The work of [1] introduced a method for splitting up tasks of agents into navigation and combat controllers and train them in separate cycles. Combinations of the controllers [2] were experimented on as well. Results show that FPS bots could learn decent navigational skills and effective stealth and aggressive style combat strategies corresponding training parameters. The paper of [3] focused on adaptive, unpredictable combat behaviour of bots and applied a dynamic reward scheme relating to damage caused by agents. They were able to create good combat strategies which stood their ground against state-machine bots.

3. Methodology

This paper applies tabular Sarsa(λ) algorithm with eligibility traces together with ε-greedy selection to FPS bots in learning. After successfully training separate navigation and combat controllers, an ultimate training is concluded by introducing a method for local environment analysis of agents. Investigation and experiments were carried out in a self-built first person shooting environment that includes all basic elements of a commercial FPS game. Four predefined map layouts were used, the Arena and Maze maps for training navigation, the Combat map for training combat and the Ultimate map for the ultimate training. It was important to properly divide navigation and combat tasks and implement map layouts specifically designed for them. Agents keep track of their own data fields (absolute location, direction, view range, health points and timestamp for shooting cooldown) and are able to move forwards or backwards, turn left or right, pick up an item and shoot their weapon. States of navigation training consist of sensors that locate nearby walls or items, whereas the closest visible enemy and its data is stored for states of combat training. Reward systems are divided as well – for navigation, small punishment is given for colliding, small reward for moving and large reward for collecting an item, while for combat, large rewards for hitting or killing a target, large punishment for being killed and small punishments for missing a shot or getting wounded. A small modification is added to the Sarsa(λ) algorithm – eligibility traces need to be reinitialized after every episode, stated by [4].

The proposed method is termed analysis of local environments of agents and it suggests an approach for data abstraction that chooses the more favourable (previously trained) controller at a given state based on an analysis (assigning numerical values) of the agent's nearby surroundings. The aim of the method is to be able to select a preferred behaviour considering a local area of the agent because map layouts may contain smaller or wider open areas as well as parts with a higher density of obstacles. Combining the proposed method with already trained navigation and combat controllers is termed ultimate training in this paper.

4. Results of Navigation and Combat

Recreation of navigation and combat controllers from referenced work is the base of this paper. Test runs include separate training of controllers, each for 10 000 iterations and of four FPS bots. Trials 1 to 9 differ in values of γ and λ parameters of the learning algorithm.

In terms of counts of collisions, ~1.5 times more colliding happened in the Maze map than in the Arena map. When λ is higher, collisions decrease and distance travelled by agents is also lower in the Arena map. In the Maze map, distance increased in this case, which implies that in this environment, agents can find a better navigation strategy. Also, a higher λ corresponds to more items collected. During combat training on the Combat map, stealth behaviour can be accomplished with λ = 0.0 or 0.4, which means

agents prioritize exploring the map, and λ = 0.8 cases result in an aggressive style combat. Trials of this paper come to similar conclusions that related research did, except for the maze-like training.

5. Results of Ultimate Training

Trained controllers are combined into an ultimate training and the trials 1 to 4 are varying in size of agents' local environments (7x7, 11x11, 15x15 metres), and agents' training processes turned off or being continued through the ultimate training (overall 24 test runs for 10 000 iterations with four agents on the Ultimate map).

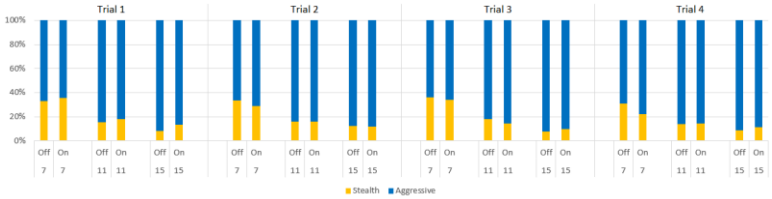

Fig. 1. Percentages of chosen behaviour of Ultimate map

Percentage results of chosen controllers through training state that there is a correspondence between preferred behaviours and local area sizes – choosing aggressive style more often as said size grows. With a smaller local environment and training turned on, choices between controllers are more balanced which would be optimal in the long run through all kinds of map layouts, also this implies a fair adaptiveness is achievable.

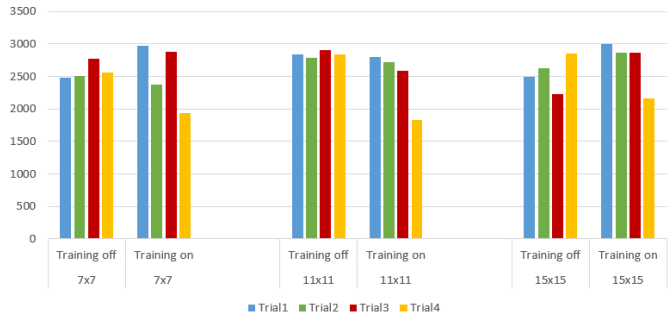

Fig. 2. Collision counts of Ultimate map

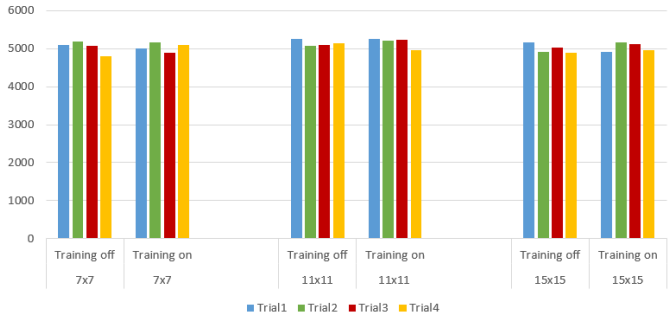

Fig. 3. Distances travelled of Ultimate map

Both in terms of collisions and travelled distances, correlations to γ and λ are observed – higher values (aggressive) indicate less colliding and distance, smaller values (stealth) result in the highest colliding and more distance travelled. These also correspond to chosen behaviours throughout the training (stealth prefers exploring the map, while aggressive does not and rather engages in combat). Also, overall distance travelled is ~1.5 times more than that of the simple navigation controller.

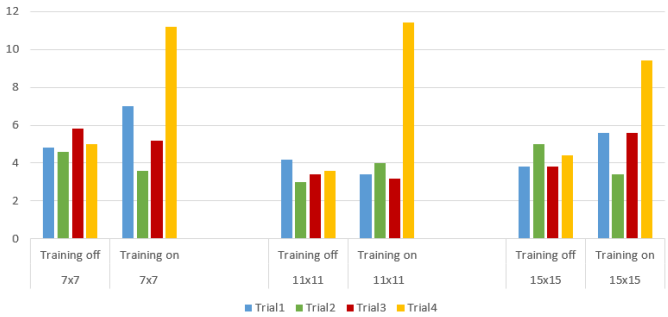

Fig. 4. Items collected of Ultimate map

There is no clear pattern of convergence of items collected to any other aspect. When training stays on, slightly more items can be picked up by agents. Trial 4 (γ and λ both set to 0.8) has a surprisingly high peak in item collection, demonstrating a newly developed combat strategy – through observation of trainings and replays, it is clear that agents realize the healing ability of the items, thus collecting more and surviving longer during combat, and such a strategy may be quite effective and useful. Other trials perform at an average level in all aspects.

Fig. 5. Numbers of kills and deaths of Ultimate map

Counts of kill and death occurrences increase (decrease) with a lower (higher) γ and λ referring to features of the stealth (aggressive) behaviour. Expect for trial 4 which has a lower kill-death count due to abilities of healing items.

6. Conclusion

Successfully trained navigation and combat controllers in an FPS game are combined as an ultimate training where a choice is made of preferred behaviour based on local environment analysis of agents. Similar levels of performance is achieved to those of related work, although in one case, a new and effective strategy is developed where FPS bots make use of the healing abilities of collectible items in order to survive longer in combat.

References

[1] Michelle McPartland and Marcus Gallagher, "Learning to be a Bot: Reinforcement Learning in Shooter Games," Artificial Intelligence Interactive Digital Entertainment, Stanford, CA, 2008.

[2] Michelle McPartland and Marcus Gallagher, "Reinforcement Learning in First Person Shooter Games," IEEE Transactions on Computational Intelligence and AI in Games, Vol. 3, No. 1, 2011.

[3] Frank G. Glavin and Michael G. Madden, "Adaptive Shooting for Bots in First Person Shooter Games Using Reinforcement Learning," IEEE Transactions on Computational Intelligence and AI in Games, Vol. 7, No. 2, 2015.

[4] John Loch and Satinder Singh, "Using Eligibility Traces to Find the Best Memoryless Policy in Partially Observable Markov Decision Processes," ICML, pp. 323-331, 1998.