University of Bristol

Dissertation

Nishant Ramakuru
1977959

Supervisor: Namid Shatil
Co-Supervisor: Dr. Sabine Hauert

First version

Bristol, September 2020

Table of Contents

1. Introduction

Abstract

Swarms of autonomous agents acting towards a common purpose have been shown to benefit when the individual agents are pre-designed to exhibit elements of cooperation and collaboration. In this project, we will extend this approach to include considerate behaviour; swarms that actively avoid inconveniencing others within the group. To achieve this, we will create swarm algorithms that include an extension to traditional game theory by conditioning each individual action on the preference of other individuals within the group. These algorithms can then be tested in a variety of scenarios, both through simulations and a swarm robotic platform.

Index: Traditional Game theory (TGT), Conditional Game Theory (CGT)

1.1 Aim

The aim of the project is showing how the use of game theory can benefit a swarm of intelligent agents and understand how robots will interact and update their strategies to align with the common goal of the swarm.

The system can be modelled at two levels; *Macroscopic*, where the entire swarm is controlled, and *Microscopic*, where each individual agent is controlled [9]. In both the methods, the interaction of the agents is not exclusively taken into consideration. Swarm algorithms rarely include the rule of taking the preferences of other agents of the swarm, or "Considerate behaviour". We define considerate behaviour as the additional rule of conditioning each agent with the actions of other agents.

The proposed method is a mixture of the two modelling scales, where the goal of the swarm will be set at the macroscopic level but an additional rule of conditioning each agent will be carried out at the microscopic level. This is different to the traditional game theory (TGT) notion of the payoff structure, as the payoff structure is designed to be conditional on the players option, so-called conditional game theory (CGT) [11]. To clearly observe the behaviour of the agents, a two agent system will be considered in this project as a starting point.

1.2 Objectives

- **Study and understand TGT**
  The first phase of the project will be to get a thorough understanding of TGT and its components.
- **Build base algorithm for TGT**
  The next step is to understand the mathematical relations and build a base algorithm using Bayesian networks and TGT models. The proposed method will focus on how to incorporate these strategies and models into a swarm algorithm.
- **Simulate TGT algorithm**
  Once the underlying concept and rules of the method are formed, the algorithms will be tested in simulated environments.The proposed method will simulate the TGT algorithm and observe the behaviour of the agents.
- **Incorporate additional rule(considerate behaviour)**
  The next step is to understand the extensions to TGT, known as CGT. Add the additional rule of conditioning each agent with the action preferences of others, using concepts from Bayes theorem. The successful execution of the algorithm will lead to a calculation of the Nash equilibrium for the swarm.
- **Simulate Considerate Behaviour**
  The proposed method will simulate the CGT algorithm and observe the functioning of the agents. The agents in the simulation should exhibit considerate behaviour and should accomplish the goal of the swarm simultaneously.
- **Reinforcement learning (benchmark)**
  An approach to the same task using reinforcement learning will also be carried out as a benchmark to test the performance of the proposed method. A reward system will be incorporated in the reinforcement learning algorithm, where the positive reward will be awarded for exhibiting properties of considerate behaviour and the accomplishment of the task. Negative rewards will be awarded if the agent somehow diminishes the performance of the other agent or does not accomplish the given task.

- **Simulate Reinforcement learning**
  The algorithm will be iterated and trained until convergence. It will be simulated in the same environment as the CGT model to normalize the parameters for comparison.

- **Hardware implementation**
  The final stage of the project will be to enable robots to exhibit the developed strategies in a physical environment, which motivate changes to the algorithm to make it more robust and pragmatic.

2. Literature Review

Swarm robotics is a system which consists of large numbers of mostly simple physical robots. The collective behaviour emerging from interactions between robots is referred to as a swarm algorithm. Game theory is the study of mathematical models of strategies based on the interactions of rational agents. The objective of this project is to implement strategies using game theory and an extension to game theory, whereby rational agents increase the performance of the group, or in this case, a swarm. Hence the prerequisites of this project require a thorough understanding of game theory, swarm algorithms and swarm simulation platforms. It also requires a basic understanding of machine learning and artificial intelligence to implement reinforcement learning as a benchmark to asses efficiency of the proposed method.

The following points briefly summarizes the literature.

- **Game theory in robotics:** Coordination vs performance shows us how important game theory can be in swarms [11] . Having intelligent agents with high individual performance with no cooperation can lead to a poorly functioning swarm. On the other hand a poor performance yielding agents with high cooperation can emerge as a good functioning swarm.

  Game theory is already being used in robotics for example, see [2] [3] which give examples of how game theory can be useful in human robot interaction, which is nothing but a two agent swarm. A more complex model, gives a example of how game theory can be used by classifying the swarms as predator and prey [1] shows us that the interpolation of game theory can be carried out in two ways. Either the agents have conflicting goals [1], or the agents work towards a common one [2]. The latter of the two, where the use of game theory is used to engineer considerate behaviour, is desired aim for proposed method.

- **Traditional (TGT):** Two classic examples of the type of scenarios that are widely used in TGT are the "Prisoner's Dilemma" [8] and "Battle Of Sexes" [5], which both refer to two agent systems. This project requires the understanding Nash equilibrium [4], which states how the strategies of the agents should be chosen in order to maximise the output of the total system of agents. This will act as the convergence measure for the algorithm.

- **Conditional Game Theory (CGT):** Conditional game theory, is an extension to traditional game theory. The underlying concept is to relate the actions of agents to the preferences of other agents belonging to the swarm. Mathematical analysis of CGT is outlined in, for example [10] [11].

- **Theoretical Ground for simulations:** Mathematical models to build simulations. Next phase of the project will be to cover CGT and prepare the ground for the theoretical simulations.

- **Simulations:** Simulations are key to this project as it will be the most time consuming and tedious task. Using simulation platforms rather than inventing from scratch will be ideal in this scenario. Platforms like ARGoS will be a good starting point to explore the features and capabilities of softwares [7]. See, for example [12] for reference to comparison of various modelling and simulation methods as per the specifications of the project.

- **Reinforcement learning:** A reinforcement learning technique implemented on a two agent system, similar to the one being proposed will be implemented to make comparison easier. Reinforcement learning in swarms with two agent system such as [6] will act as a guidance to carry out this objective. An actor critic relationship much like the predator-prey [1] relationship is implemented in this paper.

3. Methodology

Iterated Prisoner's Dilemma

A collection of strategies are implemented in an iterative based algorithm. a number of players are played against each other for generations confining to different strategies. Based on the outcome of each game i.e, played between two players, with same or different strategies, the confidence of the strategies is either increased or reduced. Using similar policies used in genetic algorithms, the stronger players are replicated and the weaker players removed.

List of Strategies

- RANDOM

  Choose a random strategy from the list of strategies.

- ALWAYS DEFECT

  Always defect no matter what the opponents decision is.

- ALWAYS COOPERATE

  Always defect no matter what the opponents decision is.

- TIT FOR TAT

  The opponents who cheated in the previous generations are stored in a list and based on this information, the player will defect or cooperate based on the opponents decisions.

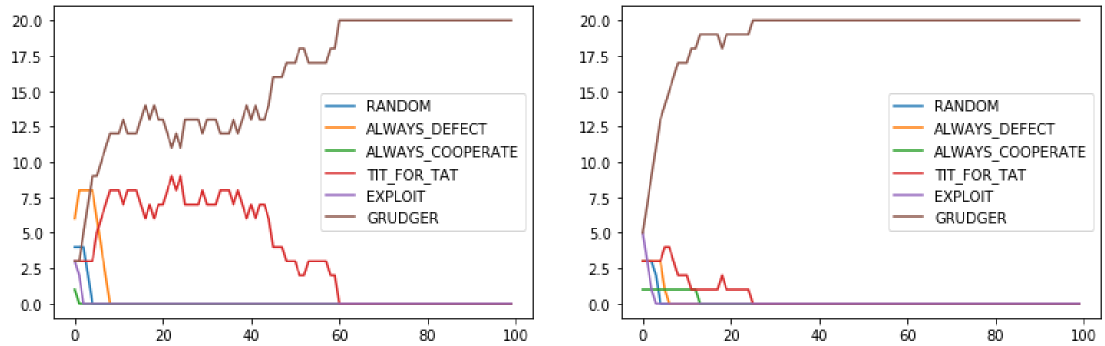
- EXPLOIT

  The opponent will exploit, ie defect if the opponent are found in the list of gullible players. A players is added to the gullible list if the player co operates if the opponent defects.

- GRUDGER

  The player will remember the opponents decions from previous generations and use that information to "hold a grudge" and choose a decision based on that information.

4. Results



(a) With limited memory **(b)** With unlimited memory

Figure 4.1: Strategy Frequencies For IDP

From 4.1 we can see that the Grudger strategy prevails in both cases.It is also observed that the grudger mirrors that of the Tit for tat strategy. In this section we discuss the effect of memory of the players.

In the case of limited memory, the players take longer to learn the best strategy.

5. Conclusion

Bibliography

[1] Haibin Duan, Pei Li, and Yaxiang Yu. A predator-prey particle swarm optimization approach to multiple ucav air combat modeled by dynamic game theory. *IEEE/CAA Journal of Automatica Sinica*, 2(1):11–18, 2015. 4

[2] Y Li, G Carboni, F Gonzalez, D Campolo, and E Burdet. Differential game theory for versatile physical human–robot interaction. *Nature Machine Intelligence*, 1(1):36–43, 2019. 4

[3] Y. Li, K. P. Tee, R. Yan, W. L. Chan, and Y. Wu. A framework of human–robot coordination based on game theory and policy iteration. *IEEE Transactions on Robotics*, 32(6):1408–1418, 2016. 4

[4] John F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950. 4

[5] Martin J Osborne and Ariel Rubinstein. *A course in game theory*. MIT press, 1994. 4

[6] J. F. Peters, C. Henry, and S. Ramanna. Reinforcement learning in swarms that learn. In *IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, pages 400–406, 2005. 4

[7] C. Pinciroli, V. Trianni, R. O'Grady, G. Pini, A. Brutschy, M. Brambilla, N. Mathews, E. Ferrante, G. Di Caro, F. Ducatelle, T. Stirling, Á. Gutiérrez, L. M. Gambardella, and M. Dorigo. Argos: A modular, multi-engine simulator for heterogeneous swarm robotics. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5027–5034, 2011. 4

[8] William Poundstone. *Prisoner's Dilemma/John von Neumann, Game Theory and the Puzzle of the Bomb*. Anchor, 1993. 4

[9] Erol Şahin, Sertan Girgin, Levent Bayindir, and Ali Emre Turgut. Swarm robotics. In *Swarm intelligence*, pages 87–100. Springer, 2008. 2

[10] Wynn C Stirling. Conditional coordination games on cyclic social influence networks. *IEEE Transactions on Computational Social Systems*, 6(2):250–267, 2019. 4

[11] Wynn C Stirling and Villino Volterra. Theory of coordinated agency. *Brigham Young University (mimeo)*, 2016. 2, 4

[12] Ying Tan and Zhong-yang Zheng. Research advance in swarm robotics. *Defence Technology*, 9(1):18–39, 2013. 4