Declaration

This report has been prepared on the basis of my own work. Where other published and unpublished source materials have been used, these have been acknowledged.

Word Count: 1,587 *2,000* intro + n *3,000* lit review + n *4,000* framework + n *4,000* experiment evaluation + n *1,000* conclusions + n *1,000* professional issues

Student Name: James King

Date of Submission: February 27, 2019

Signature:

Table of Contents

Abstract

Chapter 1: **Introduction**

1.1 Motivation

Artificial intelligence (AI) has been an idea present in the consciousness of humanity for millennia. From Hephaestus' mighty Talos to Edgar Allan Poe's commentary on 'Maelzel's Chess-Player' the idea has inspired both awe and confusion. As we move away from the mythical and the false, AI embeds itself deeper into our lives and societies. AI techniques are being used for many novel applications in areas such as medicine [9] and game playing [2].

Many of these applications are specifically using agent techniques [7, 3, 1]. Intelligent agents (IAs) have many definitions but one popular definition is that agents are anything that perceives and acts upon its environment [6]. These agents are situated in an environment, multiple of these agents can be combined in one environment to form a multi-agent system (MAS).

Within MASs it is generally possible for the agents to interact and/or communicate with each other. Communication generally occurs through agent communication languages (ACLs). The interactions that occur between agents are actions by one or all of the agents in the interaction. There is a possibility that agents can act out of pure altruism and always cooperate with other agents, however, often agents will work to protect their interests i.e. they are selfish.

It is desirable for agents in a MAS to work together to complete tasks and fulfil goals. So we want to facilitate cooperation between agents which may be selfish. There are analytical tools that we can use in game theory to understand what happens when decision-making individuals interact [4]. Some of these tools are used to explain how cooperation occurs between selfish individuals.

Many of these tools rely on the idea of reciprocal altruism [10], which stipulates that cooperation can occur between selfish individuals if they expect that cooperation to be reciprocated. Two such mechanisms are direct and indirect reciprocity. Indirect reciprocity works by an agent cooperating with another agent with the expectation that this will increase their chances of receiving cooperation from others later. Whereas direct reciprocity is expecting reciprocation of cooperation from the agent who received the cooperation in the first place.

There are many factors to consider surrounding the mechanism. One key factor is the level of visibility of the interactions within the MAS. Nowak and Sigmund [5] limit this visibility to a randomly selected group of individuals in the population (onlookers). Anothe factor is how information about the interactions can be conveyed. Sommerfeld *et al.* suggested the use of gossip [8].

So how can we use these mechanisms to facilitate cooperation between IAs? What options can we use and how can we make use of them to encourage cooperation between these agents?

1.2 Aims and Objectives

The high level aim of this project is to study how game-theoretic techniques can be leveraged in MASs to create cooperative societies of agents. To do this I have developed a theoretical framework/model using game-theoretic and MAS techniques inspired by past work in both fields. I have implemented this framework in a MAS that allows transparent decision making by agents, by which I mean agents are able to give reasons for their decisions. This implementation is on a distributable platform that allows users to set up games of the model, specify certain variables and view an analysis of the game in order to study how the mechanism and variables affect cooperation in the system.

Users are also able to create an account and refer back to games they have previously run. I have used the system to run my own experiments to study the effectiveness of the techniques I have employed and how the system needs to be set up in terms of the variables present. I have presented these experiments and then discussed the results with a conclusion on my findings and suggestions of future work as a result of these findings.

1.3 Contribution

Much of the work I have reviewed in relation to game-theoretic mechanisms has come from the field of game theory. This means the implementation of the models the authors have devised has generally not been using MAS techniques. My implementation is not only a MAS using a game-theoretic model, but it supports transparent agent decisions due to the use of the logic programming language Prolog for the implementation of agents decision making components. This implementation is also a web based system, allowing users across the world to experiment with the model I have devised.

Furthermore, the theoretical framework I have designed uses inspiration from past work on game-theoretic mechanisms but also builds upon these approaches. One way in which I build upon these is by combining the many aspects of the past approaches such as using MAS techniques to create a mixed reciprocity model (combining direct and indirect reciprocity) and using gossip as an action to convey reputation information in a mixed reciprocity model.

Lastly, in my system users associate agents with specific strategies which I have included in the system. These strategies come from past work in the topic of game theory, but are implemented using an agent architecture I have purpose designed for my theoretical framework. This included developing upon the strategies with strategy components for whatever roles agents may have at a particular timepoint in a society (such as when an agent is not in an interaction). Another development upon the current strategies are the trust models I have created for the strategies for interpreting the events in the environment.

1.4 Structure

In the next chapter of this report I have explored and described the past work that has been a central inspiration to the theoretical framework I have devised. I have included other interesting information and possible mechanisms from past work that could be used to encourage cooperation in MASs in the appendix. This background reading then leads into the next chapter in which I have defined the theoretical framework for my system.

Following on from the theoretical framework in the same chapter I have described the implementation of this theoretical framework in a web application. This application is available for users across the internet to set up games and view the happenings and analysis of their games. In the next chapter I use this platform to experiment in order to review how effective the mechanism, strategies and variables in the game are in encouraging cooperation between selfish agents.

The final two main chapters of this report is a discussion on my findings from the experiments and on the rest of the project, and then a conclusion and evoluation of the project. The appendix is at the end and contains further information referenced in the report but is not necessarily central to the project.

Bibliography

[1] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.

[2] Demis Hassabis and Silver David. Alphago zero: Learning from scratch. *deepMind official website*, 18, 2017.

[3] Özgür Kafali, Alfonso E. Romero, and Kostas Stathis. Agent-oriented activity recognition in the event calculus: an application for diabetic patients. *Computational Intelligence*, 33:899–925, August 2017.

[4] Roger B Myerson. *Game theory.* Harvard university press, 2013.

[5] Martin A. Nowak and Karl Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393:573–577, 1998.

[6] Stuart J Russell and Peter Norvig. *Artificial intelligence: a modern approach.* Malaysia; Pearson Education Limited,, 2016.

[7] Barry G Silverman, Nancy Hanrahan, Gnana Bharathy, Kim Gordon, and Dan Johnson. A systems approach to healthcare: agent-based modeling, community mental health, and population well-being. *Artificial intelligence in medicine*, 63(2):61–71, 2015.

[8] Ralf D. Sommerfeld, Hans-Jürgen Krambeck, Dirk Semmann, and Manfred Milinski. Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Sciences of the United States of America*, 104:17435–17440, 2007.

[9] Young-Geun Choi Hee-Cheon You Ja-Heon Kang Chi-Hyuck Jun Su-Dong Lee, Ji-Hyung Lee. Machine learning models based on the dimensionality reduction of standard automated perimetry data for glaucoma diagnosis. *Artificial Intelligence in Medicine*, 18, 2019.

[10] Robert L Trivers. The evolution of reciprocal altruism. *The Quarterly review of biology*, 46(1):35–57, 1971.

Chapter 2: **Appendix**

Describe the contents of my appendix.

2.1 Strategies

include list of strategy components and trust models, all associated together.