Report on indirect reciprocity, strategies for agents and the development of a concrete model to implement

James King
Supervisor: Kostas Stathis

October 2018

Abstract

Indirect reciprocity is a mechanism that uses reciprocation theory to aid in the evolution of cooperation. It is a promising motivator for cooperation in societies with agents of higher intelligence levels, such as human societies or even multi-agent systems. I plan to implement the mechanism programmatically, but there are many formulations and many possible additions. In this report I will explore past approaches to indirect reciprocity, comparing and contrasting the variations proposed and considering their impact on multi-agent systems. The outcome of this exploration of approaches will be a concrete model to implement in this project.

I. Introduction

The evolution and preservation of cooperation has been a puzzle for evolutionary theorists for a long time. Many different approaches have been taken to give an explanation to cooperative phenomena, especially from the field of game-theory. Often these approaches have come from the idea of reciprocity, where agents can grow mutually beneficial relationships by repeated interactions. The most popular mechanism of this being direct reciprocity where interactions are repeated between the same two individuals, and thus they may reciprocate directly with each other.
There is another game-theoretic mechanism known as indirect reciprocity, which works on the idea that nice agents will help those who help each other. A number of models have been proposed to run indirect reciprocity. It is these models I shall be describing and reviewing, before using them to formulate a concrete model to implement in my project.

II. Review of Past Work

i. Nowak, Sigmund and Image Scoring

I will begin by discussing possibly the most popular model of indirect reciprocity presented by Nowak and Sigmund [2].
Nowak 2005
The model, its good parts and limitations/criticisms.

ii. Standing Strategy

Leimar Hammerstein

iii. Roberts' Mixed Reciprocity Model

Gilbert Roberts compares a number of models of indirect reciprocity [3] and fuses both indirect and direct reciprocity together, to suggest a mixed framework. Roberts put forward this notion due to his perception that

indirect reciprocity alone is not a generalisable concept due to the close knit nature of many societies. In the framework Roberts puts forwards, agents base their decisions on a reputation score and/or an experience score.
The model is laid out on an island system which effectively splits the whole population into islands and then into groups on each island in which the inhabitants interact with each other. Interactions occurred by randomly selecting a donor and receiver, if the donor helps it incurs a cost of -1 and the receiver receives a benefit of 2. These interactions were repeated until on average all group members interacted with each other a certain specified amount of times.
The model includes a reproductive system where individuals reproduce if they have a measure of success locally and globally - each individual has a higher chance of reproducing locally and a smaller chance of reproducing locally. Mutation can occur with a very small chance, where a strategy is replaced by another randomly. Interestingly Roberts also considers the case where an agent has no resources to cooperate and thus must defect.

iv. Phelps' Mixed Reciprocity Model

v. Gossip and Onlookers

[4]

Simpson altruism reciprocity
Judgement Bias
Competitive altruism

III. A Concrete Model

i. Comparison of Models

The aim of Roberts [3] is to compare the effectiveness of indirect reciprocity models and direct reciprocity. Roberts' results support Nowak and Sigmund's 1998 result that image scoring can support cooperation is robust even when criticisms relating to genetic drift and errors are taken into consideration.
However, it is highlighted that image scoring does not take into account who the donor is defecting against, whereas standing strategy does. Due to this image scoring suffers from an inability to distinguish between pure defectors and those who would cooperate with a cooperator. From this paper it appears that standing strategy is not only stable with respect to other indirect reciprocity strategies and non-cooperation but also against direct reciprocity. Standing strategy gives information both about how agents have interacted and the context of this action, an advantage over image scoring.
Roberts also inadvertently provides a counterpoint to the effectiveness of gossip as an alternative to direct observation, stating that cooperation is only worth it when that cooperation is improving your chance of being cooperated with. Gossip would seem to be a barrier, especially as information could be distorted by other agents.
Another point is highlighted by Nowak and Sigmund [2] that appears to reduce the effectiveness of a model that uses gossip, stating that the probability of knowing the 'image' of the recipient must exceed the cost-to-benefit ratio of the altruistic act. Which Robert's notes is dependent on factors such as how many interactions are public and how reliably they are observed.
Roberts' model seems to more accurately portray interaction due to the possibility of basing decisions when re-meeting on experience rather than just reputation, whilst other models do not. An example of this would be where a human wishes to employ someone to do a job for them, say person A has a good reputation so they choose them. If person A does a good job their reputation may be maintained or even increased. However, if they do not even attempt to do a good job their reputation may be besmirched, but this exact person hiring them will not likely base their future decisions on person A's reputation but their experience with them.
A limitation of Roberts' model is it's inability

to use a long interaction history to motivate decisions, so not truly representing the more developed direct reciprocity mechanisms. This is analogous to our job story if person A had done a few good jobs for the person hiring them, and then one bad job. The hirer will likely base their future decisions on their history of decisions not the one just before. Though this may not always be the case, I feel it is important that direct reciprocity mechanisms should give the opportunity to view the whole history.

ii. Specification of the Model

Things to consider:

- Image scoring vs. standing strategy vs. gossip dependent vs. interaction history
- The case where an individual has no resources to cooperate with
- Direct observation vs gossip vs hybrid
- Pure reputation vs experience and reputation hybrid
- Context and action split

iii. Specification of Strategies to implement

IV. Discussion and Conclusion

References

[1] Vince Knight; Owen Campbell; Marc; ericss; VSN Reddy Janga; James Campbell; Karol M. Langner; T.J. Gaffney; Sourav Singh; Nikoleta; Julie Rymer; Thomas Campbell; Jason Young; MHakem; Geraint Palmer; Kristian Glass; edouardArgenson; Daniel Mancia; Martin Jones; Cameron Davidson-Pilon; alajara; Ranjini Das; Marios Zoulias; Aaron Kratz; Timothy Standen; Paul Slavin; Adam Pohl; Jochen MÃijller; Georgios Koutsovoulos; Areeb Ahmed. Axelrod: 4.3.0. `http://dx.doi.org/10.5281/zenodo.1405868`, September 2018.

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

[3] Gilbert Roberts. Evolution of direct and indirect reciprocity. *Proceedings of The Royal Society*, 275:173–179, September 2008.

[4] 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.