Student: Zihan Zhu
Tutor: Mirco Musolesi
Project: Understanding the Evolution of Cooperation in Societies of Artificial Agents
Date: 12th January 2022

Literature Review on Understanding the Evolution of Cooperation in Societies of Artificial Agents

Introduction

Cooperation and cooperative behaviour is a vital factor in the prosperity of human society and that of many other species. While modern specialist AI(Artificial Intelligence) has evolved huge adaptability and capability in accomplishing a variety of tasks, in the real world scenarios, the agents inevitably need to interact with one another in order to jointly improve their welfare, from daily missions such as driving on roads, scheduling meetings, to global affairs such as commerce and negotiation. As suggested the “society of minds” paradigm by Marvin Minsky[1], natural intelligence arises from the interactions of numerous simple agents, and the power of intelligence stems from the vast diversity of individuals. To take a step further the intelligence of artificial agents so that they can work collaboratively with their peers of diverse and complementary skills and aid humans in a broader scope, it will be important to equip them with the capabilities necessary to cooperate and to foster cooperation, in similar ways nature intelligence do.

This article will critically review the prior work in terms of cooperation, cooperative AI, and cooperative MAS(Multi-Agent System), including important concepts and terminologies that define cooperation in general, useful models and tournaments for studying cooperative MAS, relevant learning strategies and algorithms, and related applications. At the end of the article will be suggested some major challenges worth looking into and where the subject will be possibly trending.

1. Background of Cooperation

1.1 What is Cooperation

Cooperation can occur in situations where there are two or more individuals of their own interests. A cooperator is someone who pays a cost, c, for another individual to receive a benefit, b.[2] In the field of artificial intelligence, Cooperation can happen when agents interact with each other. A multi-agent system with agents of cooperative capability is called a Cooperative MAS. The study of such agents and systems are termed in 2020 by DeepMind the Cooperative AI[3], which includes all AI research trying to help individuals, humans, and machines, to find ways to improve their joint welfare.

1.2 Where Cooperation can Happen

The dynamics of multi-agent interaction are primarily dependent on the association between agents' payoffs.[4] There are three situations how multiple agents can be relevant to one another: when they share pure common interest, when they have mix motives, and when they have pure conflicting interest.[3] Agents are likely to cooperate in the first two situations with the presence of common interests. In reality, pure-conflict situations are rarely seen, even in zero-sum competitions there can be common interests if more than two players are involved.

1.3 Incentives for Cooperation

From an evolutionary view, there are two main incentives why two individuals tend to cooperate.

1) Kin Selection: This is deduced from the genetical kinship theory that natural selection can favour cooperation if interactants of an altruistic act are genetic relatives.

2) Reciprocation: Cooperation can be observed in unrelated individuals. There are direct, indirect, and group reciprocation, depending on various features such as the possibility of re-encounter, the reputation of the agents, and the neighbourhood.[2]

1.4 How to Evaluate Cooperation

The success of a cooperative strategy can be evaluated based on different dimensions. In game theory, Nash equilibrium describes the situation where no better strategies can be adopted. In evolutionary biology, Evolutionarily Stable Strategy (ESS) describes a strategy that is against invasion by any mutant strategy.[5] Other useful indices are the robustness of a strategy, i.e., whether a strategy can thrive in a variegated environment of other strategies and eventually eliminate them, and the initial viability of a strategy, i.e., whether a strategy can successfully invade in an environment dominated by non-cooperative strategies[6].

2. Artificial Agent Societies: the Models and the Tournaments

Games have been important testbeds for studying multi-agent systems. Models and agents have been specially tailored so as to give researchers a better view of cooperation and a platform for exploring algorithms and learning strategies that can evoke cooperative behaviours. Here are discussed some canonical models.

2.1 Social Dilemmas

Social Dilemma was formally termed by Dawes in 1980.[7] It is widely used in studying the interactions between agents since in such models cooperation and defection are both likely to happen. Most dilemmas today retain two conditions from the initial

definition: 1) a non-cooperative course of action is tempting for each individual in that it yields superior outcomes for self, 2) if all pursue this non-cooperative course of action, all end up worse off than if all had cooperated.[8] This means decision makers need to balance between short-term self interest and long-term collective interest.

2.2 MGPD: Matrix Game Prisoner's Dilemma

MGPD is a social dilemma game whose rule is specified by a payoff matrix. In the game, two players can take either of two actions in each turn —— to cooperate or to defect, and receive a payoff according to the actions of both. The payoff matrix is designed following T > R > P > S, R > (S+T)/2:

	C	**D**
C	R,R	S,T
D	T,S	P,P

Figure 1: the two actions are cooperation and defection. The four outcomes are R, rewards for mutual cooperation, P, punishment for mutual defection, S, sucker outcome for whom cooperates with a defector, and T, temptation outcome for whom defects against a cooperator.

Early-stage studies relied largely on this model, from which fundamental theories were deduced.

2.2.1 MGPD with Reciprocation

In 'the Evolution of Cooperation'[6], experiments were done on MGPD, moreover, it novelly assumes that two individuals can meet more than once and that the number of interactions is not fixed in advance, by introducing a possibility w of a player knowing the strategy of its partner. This simulates reciprocation in the real world.

This model produced the famous TFT(TIT FOR TAT) strategy, an evolutionarily stable strategy that can foster and sustain cooperation in an environment providing that there is a fair number of interactions, i.e., w is large enough.

2.2.2 MGPD with Reciprocation and Kinship

Later research [9] took also kinship theory into account by introducing another possibility r of non-random interactions where like-individuals interact with like.

It confirmed the superiority of TFT against ALL D strategy, when the 'doses' of reciprocity and kin selection went beyond certain threshold.

2.2.3 Discussion

Researches based on MGPD yielded great success in early phase by producing the famous TFT strategy and revealing the close relation between cooperation and kinship selection and reciprocation. However, the model has several deficiencies. Firstly, it

is limited by simple rules thus fails to simulate complicated real-world scenarios that are temporally and spatially extended. Secondly, it treats cooperation and defection as binary choices which in reality are more like graded properties. Early paper selected strategies from pools of hand-coded algorithms only and used evolutionary computation approaches, thus suffered from their drawbacks.

2.3 Multi-agent Models

Modern study of cooperation in AI area requires more expressive tournaments and powerful agents. Models have been invented where machine learning techniques can be utilized to find out what environment, what agents, and what strategies will foster cooperation.

2.3.1 Background: Markov Game

Markov game is a more expressive, flexible, and rigorous framework suitable for studying complicated MAS problems. Such models can be temporally extended and partially observable to players, thus better capture real-world situations than MGPD. It is normally defined by a set of states and actions, a transition function to transit states according to agents' actions, and a reward function.[10] In multi-agent systems, agents usually share common states, but each has a unique set of actions and reward function.

Many learning strategies, such as Q-learning, allow agents to take actions that maximize their discounted reward, i.e., long-term payoff of actions. Here are discussed some valuable research outcomes based on such models.

2.3.2 Partner Selection: Agents of Cooperative Capability

[11] defined an N-agent Dilemma game where each agents is endowed with the capability of partner selection, i.e., the freedom to choose its partner every round. In the training, agents not only learn to play, but are given the interaction history of all other agents, based on which they learn to choose partners. The result showed that this capability could promote cooperative behaviours even though agents were trained to maximize a purely selfish objective function in a decentralized environment. and it was a bottom-up approach that resulted in the well-known strategy, TFT. The underlying reason is that the agents quickly learn to play with those who have cooperated before, then learn to cooperate so that they can be chosen by others and be given the opportunity to potentially receive rewards. TFT is eventually evolved instead of ALL C because it can effectively regulate the exploitative behavior of defecting agents.

However, the conclusion was drawn subject to certain environment variables, such as reward and the amount of information available to each agent, so the scalability is in doubt. The paper managed to solve non-stationarity issue in RL(Reinforcement Learning)

by refreshing memory buffer every episode, which, though worked well, cannot guarantee that the agents are responsive enough to the most fresh information.

2.3.3 Sequential Social Dilemma: Cooperative Parameters

[12] designed a SSD model comprising a Markov structure and two disjoint sets of policies demonstrating cooperation and defection respectively. In this setting, cooperation and defection are reflected not on a sole action but on the overall policy agents adopt, measured by certain threshold value. From this, two dilemma games are designed in the paper to study what parameters influence agents' willing to cooperate.

- Environment: the 'Gathering' game showed that resource abundance and conflict-cost in an environment influence the aggressiveness of agents. Cooperation is more likely to emerge when resources are plentiful, and vice versa.
- Intrinsic property of agents: the parameters in the learning algorithm influence how cooperative the agents are. Those of greater discount parameter have better memory, thereby more readily to defect. Those of large batch size are more experienced, thereby can take evasive actions from being tagged while cooperating. Those of more hidden units in neural network are interpreted to have higher cognitive capacity, thereby more likely to develop behaviors against the game's default.
- Model: SSD can be cooperate-default or defect-default, e.g., in a cooperate-default model, cooperative policies are more easy-to-learn. Agents may learn differently depending on what model is used despite that the learning algorithms are the same.

The general method in the paper of tracking social behavior metrics in addition to reward while manipulating parameters of the learning environment is widely applicable.

The paper also associated the influential parameters with factors identified by social psychology, and tried to explain the former from the later's view. However, there was no direct evidence to support such deductions and associations. Also, the results were produced with few trials, e.g., they only compared batch sizes of 1e5 and 1e6. More controlled experiments are required to draw a convictive final conclusion.

2.3.4 A Model with Reciprocation

Reciprocation can foster cooperation in MGPD, [13] showed that it is also the case in more complicated environment using RL. In the paper, two types of agents were designed: the innovators who learn purely from the environment reward, and the imitators who learn to match the sociality level of innovators. Two variants of imitators were implemented for comparison, one using a hand-coded metrics to measure sociality, the other using 'niceness network' by learning online. The agents were devoted to three dilemma games, a two-player coin dilemma, a multi-player public goods dilemma, and a

multi-player commons dilemma. Results showed that imitators outperformed the greedy baseline in all three games, and outperformed TFT in complex multi-player environments. The reciprocating agents demonstrate an ability to elicit cooperation in otherwise selfish individuals even better than those of TFT policy, this is because the former produces very clear responses to defection, thus provides a better RL signal driving innovators towards pro-sociality.

Though the model is easily scalable to complex environments, it is limited to symmetric models due to the mechanism of imitators. Also, the paper used two mutually exclusive imitators, further work can be done to combine metric-matching with niceness network for creating more pro-social agents.

3. How to Study the Agents: the Methods and the Algorithms

In this section will be reviewed methodologies and algorithms used in the study of cooperation in AI societies. We focus most on the fruits from the area of multi-agent systems. Due to the complexity of the problems dominated by the number of agents involved and the dynamics of their behaviours, rule-based methods are barely helpful. Therefore, most studies have been adopting machine learning techniques and or stochastic search methods.

3.1 Evolutionary Computation

Early studies of the evolution of cooperation were hugely inspired by biological evolution and Darwin's theory. In evolutionary computation, an initial set of candidate solutions is generated and iteratively updated. Each new generation is produced by stochastically removing less desired solutions, and introducing small random changes. In this way, researchers can eventually yield highly optimized strategies that have 'survived' from the selection process.

Arguably, this method is likely to produce cooperative strategies that intuitively resemble the process how cooperation occurs in nature. They have worked well in practice and given birth to some fundamental concepts and norms of cooperation. However, the algorithm itself cannot generate any strategy, but merely select them. It can only produce a solution that is superior to its peers, but not the optimal one among all.

3.2 Reinforcement learning

Among three main machine learning techniques, supervised learning requires providing the correct output, which is not suitable since correct polices can not be known beforehand. In unsupervised learning, no feedback is provided at all, which does not fit either since some expectations can actually be provided. In practice, RL methods are the

most favourable ones in studying cooperative MAS problems. It is inspired by the concept of dynamic programming and often modelled as a Markov Decision Process as discussed in 2.3.1.[14] The learning process enables agents to choose optimal actions by experiencing the consequences without requiring them to build full maps of the model. Popular methods include Deep Q-learning[15] and A3C Algorithm[16], and useful techniques such as epsilon-greedy action selection.

Though many RL methods have convergence proofs, convergence is not guaranteed in real-world applications.[17] It should be noted the success of training counts hugely on how well the algorithm and the model are designed. Main difficulties in applying RL to cooperative problems are credit assignment, non-stationarity, and incentive misalignment.[11] New algorithms are urgently needed to tackle those issues.

3.2.1 Team Learning

In team learning, a single learner is discovering a set of behaviours for all agents in the system. This method is easy to implement since it considers the whole team as an entity, thus avoids inter-agent credit assignment and non-stationarity issues. However, team learning is criticized for its space complexity, i.e., the explosion in the state space with multiple agents. It also does not fit in models whose domains are of inherently distributed data.

3.2.2 Concurrent Learning

As opposed to team learning, in concurrent learning, each agent has its own learning process that improves parts of the team. Unlike team learning, concurrent learning divides a large population into separate individuals, which dramatically shrinks the search space and thereby reduces computational complexity and increases flexibility. However, problems arise from the fact that the behaviours adapted by the agents based on the environment in turn affect the state of the environment, i.e., non-stationarity, which leads to the violation of some basic assumptions behind traditional machine learning. Possible solutions include modelling opponents as part of the environment, combining joint-action learning and importance sampling methods. Credit assignment is also difficult where researchers should decide between global and local rewarding policies[17].

4. Applications

Applications of Cooperative AI can be found in studying other relevant subjects and solving real-world issues.

4.1 Theoretical Support

The models and concepts developed from this topic can be adopted in other subjects and communities working on cooperation, including in the natural, social, and behavioural sciences. For example, the models and methods can be used to analyse games and strategies in game theory. Multi-agent systems can serve as frameworks for studying cooperative phenomenons within and between species in biology, chronic and acute diseases in medical science, and cooperative human behaviours in social sciences.

4.2 Robotics, Games, and Sports

Cooperative multi-agent systems can be applied in industry so that specialist agents of complementary skills can develop coordinated behaviours and work together for complicated tasks. In fields of sports and games, the agents can help coaches and professionals practice and explore tactics both in simulation and with real robots. In games such as MOBA games, agents with more powerful cooperative capabilities can serve as practicing partners or challenges for human players and enhance their game experience.

4.3 Daily-life Application

Cooperative AI will aid humans in their daily life. It can provide solutions to numerous problems of cooperation, such as vehicle monitoring, meeting scheduling, air traffic control, network management and routing. It will also contribute to solving international issues such as pandemic preparedness and disarmament.

5. Conclusion

Cooperation plays an important role in the prosperity of human society and is crucial for human well-being. The capability of cooperation is essential in solving a broad range of complicated problems in the real world.

In the field of artificial intelligence, concepts about cooperation are serving as the foundation for studying cooperative systems and behaviours. Advances in algorithms and research methods are providing new scientific tools for understanding such systems and devising novel cooperative structures. Research outcomes in AI are being applied in a way that benefits human and providing viewpoints for other related fields.

The study of cooperative AI and MAS is still in its infancy. There are several directions that worth further study:

- *Scalability*

Many researches were focusing on simple Markov models and merely two-agent systems, which certainly fail to simulate real-world situations. It is suggested that models

should be able to scale up by the number of agents involved, the diversity of agents in their capabilities and goals, and agents' internal or hidden states.[18]

• *Adaptive dynamics and Nash equilibria*

Multi-agent systems are hard in terms of their dynamic property. Better learning methods and algorithms should be explored for agents to better understand their surroundings and in finding effectively the Nash equilibria or even best policies that promote cooperation[19]. The algorithms should also be adaptive to dynamic changes that introduce new agents, abilities, or goals into a running system.

• *AI systems that interact with humans*

Most researches are focusing on cooperation between agents. In the future, AI will need to cooperate with not only other AI, but also humans. To make the interactions between AI and humans more natural and fluent, we should endow it with some humanoid characteristics, such as the abilities of reasoning and modelling humans' mental states to understand social norms and support complex teamwork[20].

In conclusion, studies in narrow AI have achieved significant success. Artificial intelligence has gained abilities in vision, natural language, decision making and etc, as good as human using their perceptions and brain. Now It is time to examine the possibility of enabling artificial intelligence to cooperate and work altogether as a coherent system. This will be vital in advancing AI to a level where it may in one day evolve to a comprehensive entity the rivals human.

Reference:

[1] M. Minsky, 'The Society of Mind', Simon and Schuster, Mar. 1988.

[2] M. A. Nowak, 'Five Rules for the Evolution of Cooperation', *Science*, vol. 314, no. 5805, pp. 1560–1563, Dec. 2006, doi: 10.1126/science.1133755.

[3] A. Dafoe *et al.*, 'Open Problems in Cooperative AI', *arXiv:2012.08630 [cs]*, Dec. 2020. Available: http://arxiv.org/abs/2012.08630.

[4] S. S. Komorita and C. D. Parks, 'Interpersonal Relations: Mixed-Motive Interaction', *Annual Review of Psychology*, Vol. 46:183-207, Feb. 1995, doi: 10.1146/annurev.ps.46.020195.001151.

[5] J. M. Smith, 'Evolution and the Theory of Games: In situations characterized by conflict of interest, the best strategy to adopt depends on what others are doing', *American Scientist*, vol. 64, no. 1, pp. 41–45, 1976.

[6] R. Axelrod and W. D. Hamilton, 'The Evolution of Cooperation', *Science*, vol. 211, pp. 1390-1396, Mar. 1981, doi: 10.1126/science.7466396.

[7] R. M. Dawes, 'Social dilemmas', *Annual Review of Psychology*, vol. 31:169–193, Feb. 1980, doi: 10.1146/annurev.ps.31.020180.001125.

[8]
P. A. M. Van Lange, J. Joireman, C. D. Parks, and E. Van Dijk, 'The psychology of social dilemmas: A review', *Organizational Behavior and Human Decision Processes*, vol. 120, no. 2, pp. 125–141, Mar. 2013, doi: 10.1016/j.obhdp.2012.11.003.

[9]
S. B. Ale, J. S. Brown, and A. T. Sullivan, 'Evolution of Cooperation: Combining Kin Selection and Reciprocal Altruism into Matrix Games with Social Dilemmas', *PLoS ONE*, vol. 8, no. 5, p. e63761, May 2013, doi: 10.1371/journal.pone.0063761.

[10]
M. L. Littman, 'Markov games as a framework for multi-agent reinforcement learning', *Machine Learning Proceedings 1994*, p. 157-163, doi: 10.1016/B978-1-55860-335-6.50027-1.

[11]
N. Anastassacos, S. Hailes, and M. Musolesi, 'Partner Selection for the Emergence of Cooperation in Multi-Agent Systems Using Reinforcement Learning', *arXiv:1902.03185 [cs]*, Nov. 2019. Available: http://arxiv.org/abs/1902.03185.

[12]
J. Z. Leibo, V. Zambaldi, M. Lanctot, J. Marecki, and T. Graepel, 'Multi-agent Reinforcement Learning in Sequential Social Dilemmas', *arXiv:1702.03037 [cs]*, Feb. 2017. Available: http://arxiv.org/abs/1702.03037.

[13]
T. Eccles, E. Hughes, J. Kramár, S. Wheelwright, and J. Z. Leibo, 'Learning Reciprocity in Complex Sequential Social Dilemmas', *arXiv:1903.08082 [cs]*, Mar. 2019. Available: http://arxiv.org/abs/1903.08082.

[14]
D. Hadfield-Menell, S. J. Russell, P. Abbeel, and A. Dragan, 'Cooperative Inverse Reinforcement Learning', *arXiv:1606.03137v3 [cs.AI]*, Nov. 2016. Available: https://arxiv.org/abs/1606.03137.

[15]
C. J.C.H Watkins, 'Technical Note Q-learning', *Machine Learning*, 8, 279-292, Kluwer Academic Publisher, 1992.

[16]
V. Mnih *et al.*, 'Asynchronous Methods for Deep Reinforcement Learning', *arXiv:1602.01783 [cs]*, Jun. 2016. Available: http://arxiv.org/abs/1602.01783.

[17]
L. Panait and S. Luke, 'Cooperative Multi-Agent Learning: The State of the Art', *Auton Agent Multi-Agent Syst*, vol. 11, no. 3, pp. 387–434, Nov. 2005, doi: 10.1007/s10458-005-2631-2.

[18]
L. Busoniu, R. Babuska, and B. De Schutter, 'A Comprehensive Survey of Multiagent Reinforcement Learning', *IEEE Trans. Syst., Man, Cybern. C*, vol. 38, no. 2, pp. 156–172, Mar. 2008, doi: 10.1109/TSMCC.2007.913919.

[19]
Y. Shoham, R. Powers, and T. Grenager, 'Multi-Agent Reinforcement Learning: a critical survey', 2003.

[20]
E. Bertino and D. Lopresti, 'Artificial Intelligence & Cooperation', *arXiv:2012.06032 [cs.CY], Dec, 2020*. Available: https://arxiv.org/abs/2012.06034.