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. Therefore, cooperation, cooperative multi-agent systems(MAS) particularly, has come into the sight of AI researchers in the past two decades. 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 could 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, including important concepts and terminologies that define cooperation and cooperative AI in general, useful models and tournaments for studying cooperative agents, 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. Cooperation & Cooperative Agents: the Definitions

1.1 What is Cooperation

Cooperation can occur in situations where there are two or more individuals each of their own self-interest. When one's behaviour results in gain in the self-interest of another, we say cooperation has occurred. In the field of artificial intelligence, artificial agents usually work for completing certain tasks. While doing so, if they at any point interact with one another where the interaction results in an increase in both the interest of the other and (whether immediate or delayed) that of their own, we say the interactants have cooperated. A multi-agent system with such agents of cooperative capability is called a Cooperative Multi-Agent System(CMAS). The study of such agents and such systems was specifically termed in 2020 by DeepMind the Cooperative AI[2], which refers to AI research trying to help individuals, humans, and machines, to find ways to improve their joint welfare.

1.2 Where Cooperation can Happen

There have long been researches on areas such as game theory, social sciences, and AI showing that the dynamics of multi-agent interaction are primarily dependent on the association between agents' payoffs.[3] There are three situations how multiple agents can be relevant to one another in an interaction in terms of their interests:

1. They share pure common interests.

This means the increase in the payoff of one agent will definitely lead to that of another. In this case, it is highly possible that agents seek cooperation with one another for the maximization of their interests.

2. They have mixed motives.

The concept of mixed-motive conflict was first introduced by economist Thomas Schelling[4] to refer to situations in which there are common interests as well as conflict between adversaries. By Robinson and Goforth, the vast majority of games are designed to be mixed-motive, i.e., non-zero-sum games.[5] In this case, agents can still choose to cooperate adopting certain strategies. Those agents whose actions benefit others despite losses to their own are sometimes termed altruists, while those conducting narrow self-maximization behaviours are called egoists[6]. Non-zero-sum models coincide with human and animal interactions thus the study in these games will be particularly fruitful.

3. They have pure conflict interests.

This is when agents are playing zero-sum games, where the increase of one's payoff certainly leads to the decrease of some others. Early-phase study in biological evolution, following Darwin's theory, was based on this model and about 'minimax' strategies and the Nash equilibria. Pure conflict games de facto are rarely seen in real-world scenarios except for two-player games like chess or go. Remarkably, As the number of players increases, players can share common interests despite that the game is zero-sum.

1.3 Incentives for Cooperation

It is crucial to know what features particularly foster cooperation since it will play a significant role in the design of the algorithms, specifically, the reward functions if machine learning techniques are to be applied. Researches have drawn two main incentives why individuals tend to cooperate, which suit both in nature and in virtual agent societies.

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. Relatedness can be exhibited in Artificial agents as the probability of sharing a gene between any two participants.[7]

2. Reciprocation

Cooperation can also be observed in unrelated individuals. Reciprocal cooperation can be subdivided into categories depending on various features, such as the possibility of re-encounter, the reputation of the agents, and the neighbourhood.[8]

1.4 How to Evaluate Cooperation

The success of a cooperative strategy can be evaluated based on different dimensions. A good evaluation standard helps select superior strategies and serves as a useful reference to machine learning methods. One famous concept that originated from biological evolution is Evolutionarily Stable Strategy (ESS). It describes a strategy that is against invasion by any mutant strategy, i.e.,

A strategy I is ESS if for all mutant strategies J,

EI(I) ≥ EJ(J) and EJ(I) > EI(J),

where EJ(I) means the payoff to I playing against J.

Notably, an ESS is an equilibrium refinement of the Nash equilibrium, being already a Nash equilibrium that is also "evolutionarily stable”.[9]

Other useful indices are 1) the robustness of a strategy, i.e., whether a strategy can thrive in a variegated environment of more or less sophisticated strategies and eventually eliminate them. 2) the initial viability of a strategy, i.e., whether a strategy can successfully invade in an environment dominated by non-cooperative strategies.

2. Artificial Agent Societies: the Models and the Tournaments

Games have been important testbeds for studying cooperative behaviours of artificial agents. Most of them are specially tailored models in forms of mixed-motive games where cooperation and competition are both likely to happen. Such experiments will give researchers a better view of cooperation and a platform for exploring algorithm advances. Here are discussed comparatively some canonical models and tournaments.

2.1 Social Dilemmas

Social Dilemma was formally termed by Dawes in 1980.[10] 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.[11] Many games suitable for the study of cooperative AI are designed based on this model, such as the famous Prisoner’s dilemma, Chicken, and Assurance Dilemma.

2.2 Matrix Games

Matrix games are those whose rules are specified by a payoff matrix. To set up a Social Dilemma tournament in matrix-game form, normally, players can take either of two actions —— to cooperate or to defect. Depending on the actions they have taken, they will have four different outcomes. The matrix for a two-player social dilemma is illustrated in Figure 1.

	C	**D**
C	R	S
D	T	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. The matrix is interpreted that the payoff to C against D is S.

Matrix games are intuitive and easy to operate, so serve as useful tools for the early-stage study of cooperation. However, they fail to describe sophisticated structures.

2.2.1 Prisoner’s Dilemma in Matrix

Prisoner’s Dilemma is a classic dilemma model for studying cooperative behaviours. The payoff matrix is designed following $T > R > P > S, R > (S+T)/2$:

	C	**D**
C	b-c	-c
D	b	0

Figure 2: a benefit-cost payoff matrix of Prisoner’s dilemma, b for benefit, c for cost, where b > c, b, c $\in$ N+.

In the paper ‘the Evolution of Cooperation’ [12], the model 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.

Later researches took also kinship theory into account by introducing another possibility r of non-random interactions where like-individuals interact with like.[13]

Experts from different areas have been attempting to explore the best strategies for this dilemma. There are two ESSs worth mentioning.

1. ALL D strategy:

   To always defect regardless of what the other player does.

2. TIT FOR TAT strategy:

   To cooperate the first time, then play whatever the other player has played in the previous turn.

It is clear to see that ALL D strategy is evolutionarily stable regardless of the number of interactions between the same two individuals since the other player cannot do any better whether they choose to cooperate or defect.

On the other hand, TIT FOR TAT is an ESS provided that there is a fair number of interactions, i.e.,$w > (c - rb)/b(1 - r)$, between the two individuals, as shown in figure 3. It has also been proven a strategy of initiation viability and robustness. The discovery of the success of this strategy is especially meaningful since it can foster and promote cooperation in a population.

	TFT	ALL D
TFT	r(b-c) + (1-r)(b-c)	r(b-c) - (1-r)(1-w)c
ALL D	r(0) + (1-r)(1-w)b	0

Figure 3: payoff matrix of iterated, non-random Prisoner’s dilemma of TFT versus ALL D.

2.4 Markov Games

Markov games are of a larger scope that is inclusive of the matrix game social dilemmas. It is more expressive and powerful. Markov models can be rigorously defined and serve as a framework for studying complicated MAS problems.

2.4.1 Background

Markov Decision Process (MDP) is defined by

A set of states S and actions A.

A transition function $T : S \times A \rightarrow \Delta(S)$, where Δ(S) represents the set of discrete probability distributions over S. It defines what effects the actions will bring to the state of the environment.

A reward function $R : S \times A \rightarrow \mathbf{R}$ specifying the expectation of the agent.

Generally speaking, the agent is expected to make a decision on the current action provided with its interaction history, so as to yield the maximum discounted reward. The discounted reward evaluates long-time payoff, defined by

$$\mathbb{E}_{s_{t+1} \sim T(s_t, a_t)}[\sum_{t=0}^{\infty} \gamma^t R_t(s_t, a_t)]$$

Where t is the time stamp, Rt is the reward received at time t following the reward function, $\gamma \in [0,1]$ is the discount factor confining how much effect future rewards can bring to the current decision.

2.4.2 Multi-player Markov Game

The MDP framework is often used as a superior alternative to iterated matrix games with multiple players involved.[14] Specifically, discrete-time partially observable Markov process is preferred[15], e.g., in modelling sequential social dilemmas. Such a model is defined by

A set of states S.

A collection of sets of actions$A1,A2,\ldots,Ak$, one for each agent.

A transition function $T : S \times A_1 \times A_2 \times \ldots \times A_k \rightarrow \Delta(S)$.

Reward function $Ri : S \times A_1 \times ... \times A_k \rightarrow \mathbf{Ri}$ for agent i.

When multiple players are involved, agents who observe only their local environment need to consider the impacts of others' possible movements. In that sense, we further introduce

An observation possibility function $O : S \times 1,2,...,k \rightarrow R^d$, which states an agent's local environment in a given state.

Observation space $O_i = \{o_i \,|\, s \in S, o_i = O(s,i)\}$ for agent i, which includes a sequence of observations available to the agent.

Policy $\pi_i : O_i \rightarrow \Delta(A_i)$ for agent i to choose actions based on their observation.[16]

Following this model, a two-player PD(Prisoner's Dilemma), for example, can be translated to a Markov Process.[17] We define the actions $A1,A2 = \{C,D\}$, The payoff to player i playing a policy π1 against the other playing π2 from the initial state $s_0 \in S$ is therefore

$$V_i^{\vec{\pi}=(\pi_1,\pi_2)} = \mathbb{E}_{\vec{a_t}\sim\vec{\pi}(O(s_t)),s_{t+1}\sim T(s_t,\vec{a_t})}[\sum_{t=0}^{\infty} \gamma^t R_{it}(s_t, \vec{a_t})]$$

The four outcomes are then

$$R(s) := V_1^{\pi^C,\pi^C}(s) = V_2^{\pi^C,\pi^C}(s),$$

$$P(s) := V_1^{\pi^D,\pi^D}(s) = V_2^{\pi^D,\pi^D}(s),$$

$$S(s) := V_1^{\pi^C,\pi^D}(s) = V_2^{\pi^D,\pi^C}(s),$$

$$R(s) := V_1^{\pi^D,\pi^C}(s) = V_2^{\pi^C,\pi^D}(s).$$

The model can express other games with minor changes in the definition, for example, when studying cooperative games like Hanabi, a set of possible joint goals G can be introduced to the rewarding function as R: $G \times S \times A \rightarrow$ R. This will encourage the agents to take actions that are beneficial to the collective.

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

In this section will be reviewed general methodologies and algorithms that have been 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, i.e., survival of the fittest in natural selection. The algorithms developed based on those theories, whether heuristically or experimentally, in computer science, are termed evolutionary computation. 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 also in practice and given birth to some fundamental concepts and norms for cooperation. However, one drawback is that any evolutionary algorithm only produces a solution that is superior to its peers, 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 strategies 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, Reinforcement Learning methods, Q-Learning especially[18], 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 2.4.[19] The learning process enables agents to choose optimal actions by experiencing the consequences without requiring them to build full maps of the model.

Q-learning and other reinforcement methods have convergence proofs. Unfortunately, convergence is not guaranteed in real-world applications.[20] It should be noted the success of the learning counts hugely on how well the model is designed. The main difficulties in applying RL to cooperative problems are credit assignment, non-stationarity, and incentive misalignment. Agents should not only take actions that maximize their immediate rewards but also care about what the other agents learn. The consequences of defection should be appropriately reflected in a trajectory.[21] The notion of intrinsic rewards or agent preferences is often utilized during training to drive agents to learn coordinated behaviour that emphasizes cooperation[22].

3.3 Team Learning

Team learning is a method where 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 so that inter-agent credit assignment can be ignored. However, team learning is criticized for its space complexity, i.e., the explosion in the state space with multiple agents. It does not fit in models whose domains are of inherently distributed data.

3.4 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, which leads to the violation of some basic assumptions behind traditional machine learning. Credit assignment is also difficult where researchers should decide between global and local rewarding policies.

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 to 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 for devising novel cooperative structures. Research outcomes in AI are being applied in a way that benefits human and provide 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 research papers 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 scaled up by the number of agents involved, the diversity of agents in their capabilities and goals, and agents' internal or hidden states.[23]

- *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 of a given model.[24] The algorithms should also be adaptive to dynamic changes that introduce new agents, abilities, or goals in 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[25].

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 humans.

Reference:

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

[2]
A. Dafoe *et al.*, 'Open Problems in Cooperative AI', *arXiv:2012.08630 [cs]*, Dec. 2020, Accessed: Jan. 04, 2022. [Online]. Available: http://arxiv.org/abs/2012.08630

[3]
'Interpersonal Relations: Mixed-Motive Interaction', p. 25.

[4]
T. C Schelling. The Strategy of Conflict. Harvard University Press, Cambridge, MA, 1980.

[5]
D. Robinson and D. Goforth, 'The topology of the 2x2 games: A new periodic table', Jan. 2005. doi: 10.4324/9780203340271.

[6]
R. Axelrod, 'The Emergence of Cooperation among Egoists', *Am Polit Sci Rev*, vol. 75, no. 2, pp. 306–318, Jun. 1981, doi: 10.2307/1961366.

[7]
W. D. Hamilton, 'The Genetical Evolution of Social Behaviour. I', J. Theoret. Biol. 7, 1. doi: 10.1016/0022-5193(64)90038-4.

[8]
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.

[9]
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.

[10]
R. M. Dawes, ‘Social dilemmas’, Annual Review of Psychology, 31, 169–193, 1980. doi: https://doi.org/10.1146/annurev.ps.31.020180.001125

[11]
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.

[12]
R. Axelrod and W. D. Hamilton, ‘The Evolution of Cooperation’, vol. 211, p. 7, 1981.

[13]
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.

[14]
Y. Shoham, ‘Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations’, p. 532.

[15]
S. Yue, K. Yordanova, F. Krüger, T. Kirste, and Y. Zha, ‘A Decentralized Partially Observable Decision Model for Recognizing the Multiagent Goal in Simulation Systems’, *Discrete Dynamics in Nature and Society*, vol. 2016, pp. 1–15, 2016, doi: 10.1155/2016/5323121.

[16]
M. L. Littman, ‘Markov games as a framework for multi-agent reinforcement learning’, in *Machine Learning Proceedings 1994*, Elsevier, 1994, pp. 157–163. doi: 10.1016/B978-1-55860-335-6.50027-1.

[17]
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, Accessed: Jan. 04, 2022. [Online]. Available: http://arxiv.org/abs/1702.03037

[18]
C. J.C.H Watkins, ‘Technical Note Q-learning’, Machine Learning, 8, 279-292, Kluwer Academic Publisher, 1992.

[19]
D. Hadfield-Menell, S. J. Russell, P. Abbeel, and A. Dragan, ‘Cooperative Inverse Reinforcement Learning’, p. 9.

[20]
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.

[21]
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, Accessed: Jan. 03, 2022. [Online]. Available: http://arxiv.org/abs/1902.03185

[22]
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, Accessed: Jan. 04, 2022. [Online]. Available: http://arxiv.org/abs/1903.08082

[23]
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.

[24]
Y. Shoham, R. Powers, and T. Grenager, ‘Multi-Agent Reinforcement Learning: a critical survey’, p. 13.

[25]
E. Bertino and D. Lopresti, ‘Artificial Intelligence & Cooperation’, p. 4.