Multi-Agent Coordination Profiles Through State Space Perturbations

Derrik E. Asher
Information Sciences
CCDC Army Research Laboratory
Adelphi, MD, USA
derrik.e.asher.civ@mail.mil

Michael Garber-Barron
Computer Science
Cornell University
Ithaca, NY, USA
mgb249@cornell.edu

Sebastian S. Rodriguez
Computer Science
University of Illinois, Urbana-Champaign
Urbana, IL, USA
srodri44@illinois.edu

Erin Zaroukian
Information Sciences
CCDC Army Research Laboratory
Adelphi, MD, USA
erin.g.zaroukian.civ@mail.mil

Nicholas R. Waytowich
Human Sciences
CCDC Army Research Laboratory
Adelphi, MD, USA
nicholas.r.waytowich.civ@mail.mil

***Abstract*— The current work utilized a multi-agent reinforcement learning (MARL) algorithm embedded in a continuous predator-prey pursuit simulation environment to measure and evaluate coordination between cooperating agents. In this simulation environment, it is generally assumed that successful performance for cooperative agents necessarily results in the emergence of coordination, but a clear quantitative demonstration of coordination in this environment still does not exist. The current work focuses on 1) detecting emergent coordination between cooperating agents in a multi-agent predator-prey simulation environment, and 2) showing coordination profiles between cooperating agents extracted from systematic state perturbations. This work introduces a method for detecting and comparing the typically 'black-box' behavioral solutions that result from emergent coordination in multi-agent learning spatial tasks with a shared goal. Comparing coordination profiles can provide insights into overlapping patterns that define how agents learn to interact in cooperative multi-agent environments. Similarly, this approach provides an avenue for measuring and training agents to coordinate with humans. In this way, the present work looks towards understanding and creating artificial team-mates that will strive to coordinate optimally.**

Keywords—predator-prey pursuit, coordination, simulation experiments, perturbation analysis, coordination profiles

I. Introduction

Coordination in multi-agent systems is often ill-defined and only referred to anecdotally as the collaboration between agents to achieve a common goal. Likewise, analysis of such coordination generally boils down to measurements of overall task-performance as an indirect proxy for coordination. We reinterpret coordination as a group of agents in an environment or task domain with aligned goals that exhibit measurable and observable characteristics that result in effective team work. This requires agents to dynamically adjust their behaviors to account for the actions (state changes) of their partners and the environment [1, 2]. Reliable methods for measuring these characteristics or behavioral adjustments, however, have not been established.

The current paper presents a method for using perturbation strategies to measure coordination between agents in a multi-agent setting. In this work, we utilize a multi-agent reinforcement learning (MARL) algorithm embedded in a continuous predator-prey pursuit simulation environment to measure and evaluate coordination between cooperating agents (i.e., the predators) [3]. As human-agent teaming and collaborative agent training has become more prevalent, MARL has received increased attention, and in MARL simulation environments it is generally assumed that successful performance for cooperative agents necessarily results in the emergence of coordination [4, 5]. Coordination and performance are not necessarily linked, and the nature of MARL makes guaranteeing coordination between agents difficult [6, 7]. Coordination in multi-agent learning has thus far been measured as performance in tasks where success requires cooperation. When coordinated strategies are provably optimal, the task is often discretized [7], and optimality has not been shown to generalize to continuous tasks (e.g., [3, 8, 9] and here). Prior work [10-13] has demonstrated a spatial dependence between cooperating agents in a continuous environment using Convergent Cross Mapping, though a clear demonstration of coordination with the degree to which cooperative agents account for fellow cooperative teammates, still does not exist. The procedure introduced below is intended to detect and compare the typically 'black-box' behaviors that result from multi-agent coordination in spatial tasks using state space perturbations.

II. Methods

A continuous 2D multi-agent predator-prey pursuit simulation environment was used to generate coordination profiles through systematic perturbations of each agents' input. The environment consisted of three cooperative predator agents that had a shared goal of "hitting" the prey agent as frequently as possible in an episode (fixed number of timesteps). Conversely, the prey agent's goal was to evade the predators for the duration of each episode. Performance was measured as the collective number of hits that all predators made against the prey during an episode. Each predator moved at the same velocity and acceleration, whereas the prey agent had a distinct movement advantage and was able to accelerate 1.25 times faster than the predator agents up to a max speed that was 1.3 times faster than the predator agents' max speed. The simulation environment was provided by OpenAI gym [14], which was designed to aid in MARL development and testing.

A multi-agent deep deterministic policy gradient (MADDPG) algorithm was used to train all agents simultaneously [3]. MADDPG utilizes a centralized training decentralized testing regime where agents were trained using a centralized critic network with access to all agents states and

actions that allows agents to develop policies based on other agent observations and actions during training. At test time, agents were ran in a decentralized fashion where each agent's policy depends only on their local observations. [3].

Prior to perturbations, agents were trained for 100k episodes at 25 time steps per episode. The number of episodes and episode duration were based on previous work with convergent performance [10, 12, 13]. Each agent's state space contained its velocity, position, distance to other agents, and prey's velocity. The action output for an agent was acceleration. Test data was collected from two independently trained models (1000 episodes at 300 timesteps).

Perturbation approaches have been used to gain insight into the association between neural network (NN) inputs and outputs [15, 16] through manipulations of NN inputs. In multi-agent tasks where the state space between agents is explicitly linked, state space perturbations permit evaluation of the degree to which an agent's behavior impacts the actions of other agents. The current paper uses a simple variant of perturbation approaches to manipulate the state of a single agent (i.e., relative distance to other agents) to evaluate how those manipulations changed other agents' NN outputs (i.e., actions). Given a state of the environment, this specific perturbation approach shows the relative strength of influence in terms of percent change from a baseline (i.e., the non-perturbed state) that a perturbed agent has on the actions of other cooperative agents. This metric is important for the predator-prey scenario presented here where the objective is to maximize the number of prey hits, as a high hit count is not necessarily evidence that the allegedly cooperative predators are coordinating (i.e., accounting for the actions of other predators); it is possible that agents could behave selfishly and would not be impacted by perturbations to their partners' distances.

Perturbation values were bounded by the size of the environment that ranged between [-1, 1]. Values were determined by aggregating all relative distances between agents from the 1000 test episodes (baseline) per model to form the "perturbation distribution" with a median of zero that was divided into quantiles. The quantiles were chosen to symmetrically encapsulate the baseline (i.e., median at 50%). A perturbation was applied to an agent and the resulting other agents' actions were collected.

Perturbations were applied to an agent by adding the perturbation value (Model 1: +/-0.159, +/-0.085; Model 2: +/-0.174, +/-0.093) to the x or y component of its state. 300k state-action pairs (test data) were collected for all coordinating agents (three predators). Perturbations (Table 1) were applied to the x and y components of an agent's test data separately, to conservatively evaluate the impact of perturbing one dimension at a time, and likewise evaluating the impact that these single dimensional perturbations had on each of the individual dimensions of the other agents (x on x, x on y, y on x, and y on y). For brevity, the resulting coordination profiles for x and y components were combined to reduce small individual differences between the single dimension perturbation results.

To summarize, perturbed data was passed through all non-perturbed agents' NNs and new action values were generated and recorded. This process was repeated until all agents had been individually perturbed across all episodes. The median was calculated from 300k perturbations per quantile (4 perturbation values) to generate coordination profiles represented as percent change in action output from the baseline, and show how one agent's perturbations influenced the actions of others. Importantly, the measured difference in behavior provides a metric for comparing how one agent's actions depend on another agent's state. Indeed, if a predator is primarily accounting for the prey, then perturbations to another agent's state should have little to no impact on the predator's action outputs.

III. Results

To detect coordination between three cooperating agents within a multi-agent predator-prey simulation environment, a set of state-space perturbations were applied to two sets (models) of independently trained agents, and coordination profiles were ascertained (Fig. 1). If the perturbed coordination profiles reflect differences from the baseline condition, it is concluded that coordination emerged between agents.

Performance, measured as the cumulative number of hits achieved by predators in baseline episodes indicated that the 2 models did not differ significantly when implementing a 2-sample Kolmogorov-Smirnov test ($D = 0.047$, $p = 0.214$). This suggests that the two independently trained models do not meaningfully differ in performance in the absence of perturbations. Further, this indicates that the coordination profiles shown in figure 1 represent the differential effects of perturbations on the respective models, and do not reflect differences in performance in the baseline episodes.

The percent change in the respective agent's actions as a result of the perturbations per model is shown in figure 1. Percent change is computed by, 1) subtracting the baseline action value from the perturbed action value, 2) dividing by the baseline action value, 3) multiplying by 100 to determine percentage, and 4) taking the absolute value to simplify the result as a positive percent change from baseline. This calculation is performed at every timestep between each pair of agents for all perturbations. The absolute value of the median percent change in action with standard error are shown in figure 1 for each perturbed agent (A0, A1, and A2) relative to each other (Fig. 1). It is important to note that the median percent change is always 0 at the 0.50 quantile, where agents were not perturbed (i.e., perturbation value = 0).

Fig. 2 is a link-node diagram showing the relative impact agents had on each other as a result of the perturbations, summarizing figure 1. The three predators are represented as nodes, with corresponding colors for the directional arrows (links) and the relative strength of impact as the values. The values shown in figure 2 are the averaged ratios of percent change taken from figure 1. Note, a value of 1.00 would indicate that the agents had equal impact on one another (Fig. 2).

Table 1. Perturbation values per model.

Quantiles	Model 1	Model 2
0.30	-0.1586	-0.1744
0.40	-0.0845	-0.0931
0.50 (baseline)	0	0
0.60	0.0845	0.0931
0.70	0.1586	0.1744

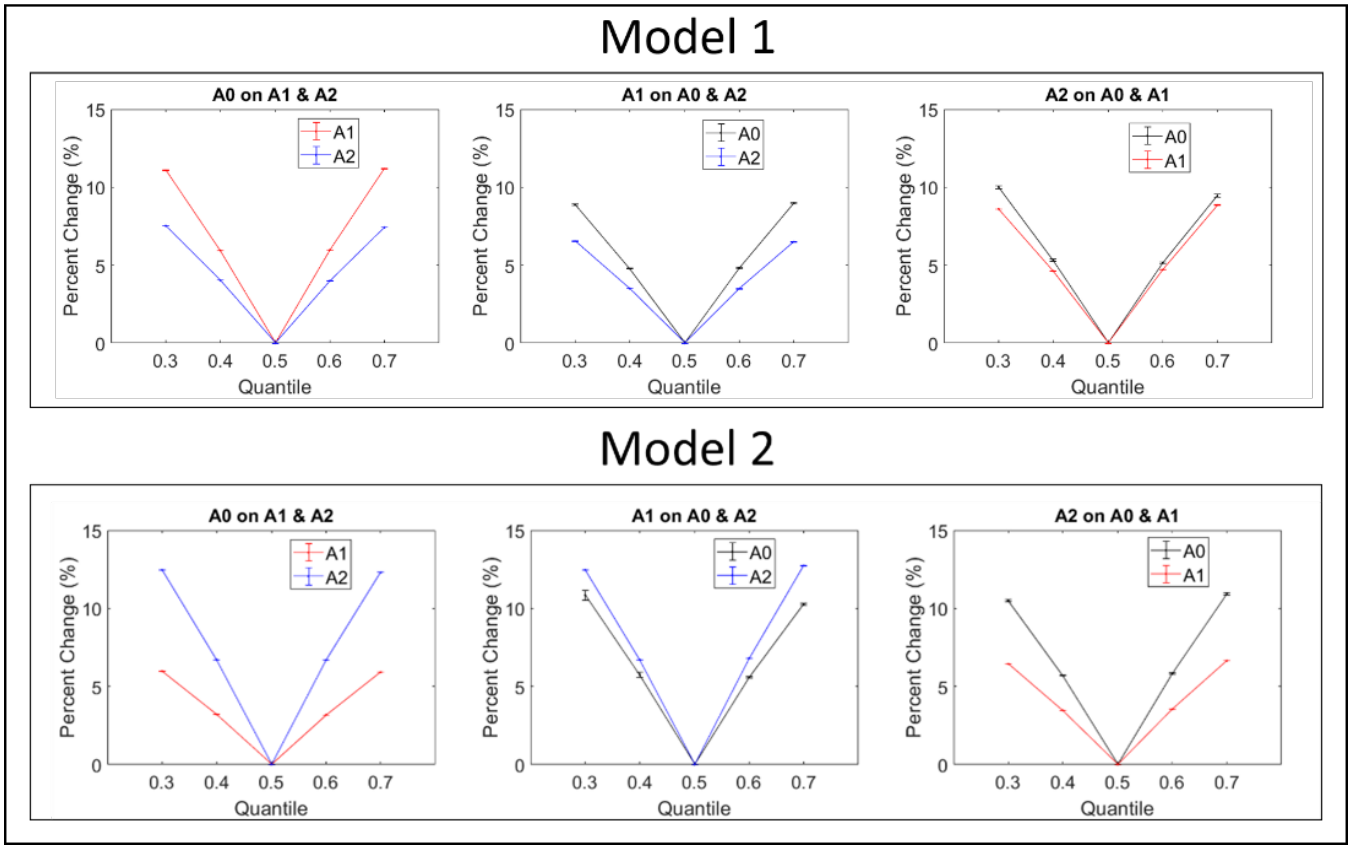

Fig. 1. Coordination profiles of percent change in action from baseline (0.50 quantile) for the effected agents. Each subplot shows the absolute value for percent change (y-axes) in action from baseline (0.50 on x-axes) for predator agents 0, 1, and 2 (A0-black, A1-red, and A2-blue). There are three subplots per model. Perturb A0 to effect A1 & A2 (left columns); perturb A1 to effect A0 & A2 (center columns); perturb A2 to effect A0 & A1 (right columns).

IV. DISCUSSION

By using a MARL algorithm embedded in a continuous predator-prey pursuit simulation environment, we recorded state spaces from two models and perturbed state space inputs (i.e., relative distances between agents) to their NNs to determine coordination dependencies between agents. This work provides the following contributions: 1) a conclusive demonstration of coordination between cooperative agents with non-zero NN output changes upon perturbations (Fig. 1, compare to baselines), 2) coordination profiles proportional to state space perturbations (e.g., Fig. 1, compare 0.70 to 0.60) that show differences between pairs of agents, and 3) a clear depiction of coordination strength with averaged percent change ratios represented as directional arrows (links) between pairs of agents (Fig. 2). The perturbation method introduced here succeeded in detecting and comparing the 'black-box' behavioral solutions that result from emergent coordination in a MARL spatial task with a shared goal.

Upon close inspection of the coordination profiles, it appears that agents may have learned to take roles within the predator-prey pursuit task. In both models, perturbation of two distinct agents led to minor changes on the third agent (Model 1: A2, Model 2: A1), while perturbation of the third agent led to much larger changes in the other agents (Fig. 2; compare arrows to and from A2 in Model 1, compare arrows to and from A1 in Model 2). The elevated percent change from baseline can be seen for the two models (Fig. 1; compare right subplot of Model 1 to middle subplot of Model 2). Further, this pattern is quite different when comparing the 2 models (roughly 1.6:1.0 for Model 1 and 3.5:1.0 for Model 2). This difference might suggest that the two models developed different coordination strategies. In the context of the predator-prey environment, this could relate to two agents taking the role of active chasers (i.e., constantly pursuing the prey), while one agent guards and tries to cut off the prey's escape route, or one agent is often passive while the other two aggressively pursue. In either case, these results suggests that the 'third' agent was less sensitive to state space information from its partners, whereas, the partners were accounting for the 'third' agents relative distance in selecting actions throughout the 1000 perturbed test episodes.

There may exist many group strategies that explain the observed coordination profiles. However, the methodology introduced here can elucidate the presence of coordination, but not the specific group strategy learned.

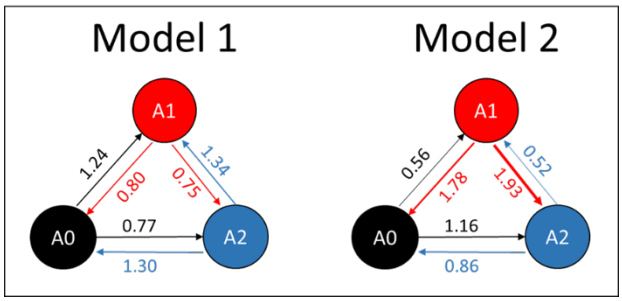

Fig. 2. Link-node diagram showing the averaged relative strength of impact as a result of perturbations between pairs of agents. Values and corresponding link thickness show the averaged ratios of percent change reflected in figure 1 between pairs of agents. Agent 0 (A0) in black, Agent 1 (A1) in red, and Agent 2 (A2) in blue.

Further, a comparison between MARL agents and human operators is warranted in future research, as we can see whether a computational solution presents the same amount of coordination and dependency to human teamwork. Being able to deter-mine the presence of coordination will allow us to integrate trained agents with humans on team-based tasks through identification of information needed to adapt to or depart from coordination, based on the needs for task completion.

ACKNOWLEDGMENT

This research was sponsored by the Army Research Laboratory and was accomplished under the Combat Capabilities Development Command Army Research Laboratory Research Associateship Program for Summer Journeyman Fellowship. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

[1] G. Klien, D. D. Woods, J. M. Bradshaw, R. R. Hoffman, and P. J. Feltovich, "Ten challenges for making automation a" team player" in joint human-agent activity," *IEEE Intelligent Systems,* vol. 19, pp. 91-95, 2004.

[2] E. Zaroukian, S. S. Rodriguez, S. L. Barton, J. A. Schaffer, B. Perelman, N. R. Waytowich*, et al.*, "Algorithmically identifying strategies in multi-agent game-theoretic environments," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, 2019, p. 1100614.

[3] R. Lowe, Y. Wu, A. Tamar, J. Harb, O. P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," in *Advances in Neural Information Processing Systems*, 2017, pp. 6382-6393.

[4] T. Engesser, T. Bolander, R. Mattmüller, and B. Nebel, "Cooperative epistemic multi-agent planning for implicit coordination," *arXiv preprint arXiv:1703.02196,* 2017.

[5] H. M. Le, Y. Yue, P. Carr, and P. Lucey, "Coordinated multi-agent imitation learning," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, 2017, pp. 1995-2003.

[6] M. Lauer and M. Riedmiller, "An algorithm for distributed reinforcement learning in cooperative multi-agent systems," in *In Proceedings of the Seventeenth International Conference on Machine Learning*, 2000.

[7] L. Matignon, G. J. Laurent, and N. Le Fort-Piat, "Independent reinforcement learners in cooperative Markov games: a survey regarding coordination problems," *Knowledge Engineering Review,* vol. 27, pp. 1-31, Mar 2012.

[8] S. L. Barton and D. Asher, "Reinforcement learning framework for collaborative agents interacting with soldiers in dynamic military contexts," in *Next-Generation Analyst VI*, 2018, p. 1065303.

[9] J. Foerster, I. A. Assael, N. de Freitas, and S. Whiteson, "Learning to communicate with deep multi-agent reinforcement learning," in *Advances in Neural Information Processing Systems*, 2016, pp. 2137-2145.

[10] D. Asher, S. Barton, E. Zaroukian, and N. Waytowich, "Effect of cooperative team size on coordination in adaptive multi-agent systems," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, 2019, p. 110060Z.

[11] S. L. Barton, N. R. Waytowich, and D. E. Asher, "Coordination-driven learning in multi-agent problem spaces," *arXiv preprint arXiv:1809.04918,* 2018.

[12] S. L. Barton, N. R. Waytowich, E. Zaroukian, and D. E. Asher, "Measuring collaborative emergent behavior in multi-agent reinforcement learning," in *International Conference on Human Systems Engineering and Design: Future Trends and Applications*, 2018, pp. 422-427.

[13] S. L. Barton, E. Zaroukian, D. E. Asher, and N. R. Waytowich, "Evaluating the Coordination of Agents in Multi-agent Reinforcement Learning," in *International Conference on Intelligent Human Systems Integration*, 2019, pp. 765-770.

[14] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang*, et al.*, "Openai gym," *arXiv preprint arXiv:1606.01540,* 2016.

[15] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "Deepfool: a simple and accurate method to fool deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2574-2582.

[16] J. Su, D. V. Vargas, and K. Sakurai, "One pixel attack for fooling deep neural networks," *IEEE Transactions on Evolutionary Computation,* 2019.