Attention manipulation in reinforcement learning agents

Oriol Corcoll (oriol.corcoll.andreu@ut.ee)
Abdullah Makkeh (makkeh@ut.ee) Jaan Aru (jaan.aru@ut.ee)
Dirk Oliver Theis (dotheis@ut.ee) Raul Vicente Zafra (raul.vicente.zafra@ut.ee)
Institute of Computer Science, Tartu University, Estonia

Abstract

The ability to change others' attention for our own benefit is referred to as attention manipulation and is known to be an important cognitive ability for coordination in cooperative tasks. In this work, we formulate attention manipulation in the context of reinforcement learning (RL) agents and argue that if the environment is complex enough agents will learn to use this skill. In particular, we first outline some of the characteristics in the environment that make it complex enough for this behavior to become relevant. Then, we test RL agents in two environments with such characteristics. Finally, we estimate a measure of attention manipulation using information theory functionals proposed to capture causal influence. Our results indicate that attention manipulation can be used by relatively simple RL agents to achieve better coordination in cooperative tasks.

Keywords: deep reinforcement learning; joint attention; attention manipulation; behavioral learning;

Attention Manipulation

Humans and other social species face the challenge of navigating through complex social interactions. In addition to the basic skills for coping with conspecifics, humans have developed specific cognitive abilities such as joint attention i.e. the ability of individuals to focus on a common goal. This ability has been proposed to be a foundation to many of our social competencies like theory of mind, it supports general cognitive development (Tomasello, 2019; Moore & Dunham, 1995). Joint attention provides the means by which humans create a joint agency with others (Bolt, Poncelet, Schultz, & Loehr, 2016) and conceive others as intentional agents (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998). An important prerequisite for joint attention is the skill of attention manipulation (Kaplan & Hafner, 2004), that is the ability to change others' behavior to attend to a shared goal.

Kaplan and Hafner (2004) provided a survey of computational models for joint attention and defined the necessary skills a robot or agent needs master to have human-like joint attention. They define as prerequisites the following: 1) attention detection, which is the ability to track the attention of others; 2) attention manipulation, use of verbal or non-verbal communication to direct the attention of others; 3) social coordination, that is, engagement in coordinated interaction with others; and 4) intentional stance, conceiving others as intentional agents. In this work, we aim to develop the skill of attention manipulation in RL agents and, by doing so, build agents with better social coordination.

Reinforcement Learning and Attention Manipulation

Deep RL has shown outstanding results in single-agent games like Chess, Go, Atari or even Starcraft. A yet challenging area for RL is the multi-agent domain, where agents need to cooperate and compete to achieve common or individual goals. In this variant, the complexity of the environment is augmented and agents have to share space, resources and goals. We argue that current reinforcement learning agents, as in the case of humans, able to manipulate the attention of others can alleviate this complexity. For agents to use this skill, environments need to be complex enough to provide the necessary pressures for this behavior to be needed. In pilot studies we have found that if the environment has the following characteristics, agents manipulate the attention of others:

- **Signalling:** in order to manipulate the attention of others, agents must be able to encode a meaningful signal that is visible to other agents. This could be achieved by incorporating a communication channel between agents or by encoding it within their movements.
- **Specialization:** agents need to be able to specialize, i.e. have particular information or skills not available to others. For example, in a foraging task, some agents may specialize on gathering fruits and others on hunting prey. However, specialization can also be achieved by limiting the information available to some agents.
- **Limited field of vision:** cooperating becomes useful when there is the need for sharing knowledge. If agents have complete vision of the environment, this need for sharing knowledge diminishes.
- **Time pressure:** attention manipulation can be used to complete tasks faster. An agent would be able to pick two boxes on its own but it would take less time if it can manipulate or alert another agent to pick one of them. For this reason, having a limited amount of resources (time, metabolism or health) adds pressure for agents to solve tasks as quickly as possible.
- **Collective tasks:** agents can have their own goals and rewards but some of this goals need to be shared with other agents, enforcing the need to coordinate their actions to achieve a common goal.

For humans, these pressures are ubiquitous in our habitat but current reinforcement learning environments do no provide them and in our view, this limits the social skills agents need to learn.

Model

The reinforcement learning policy $\pi(a_t|s_t, O_t)$ is implemented as a set of three different modules 1. The vision module extracts features from the observations generated by the environment. The attribute module processes additional attributes emitted by the environment like agent's signal state, resource's level or storage's health (see section Environments for an explanation on these attributes). Finally, the motor module will produce a movement action (left, right, up, down, stay or attack) and the expected long-term value. All these modules are implemented as fully connected layers and ReLU as activation function. The vision module has three layers, the attribute module has two layers and the motor module has one layer, all of them with 32 units. The model was trained using the Proximal Policy Optimization (PPO) algorithm (Schulman, Wolski, Dhariwal, Radford, & Klimov, 2017).

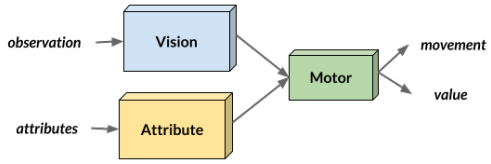

Figure 1: Architecture of the model used in the experiments. These modules are implemented with fully connected layers.

Environments

To study attention manipulation, we have created two different environments where coordination and cooperation between agents are fundamental to achieve a high collective reward. They follow the criteria described above and provide a test-bed for attention manipulation. The arena of each environment have sizes of 6x6 and 10x10, respectively. These environments produce an observation of size 3x3 and 4x4 in the first and second environments respectively. Additionally, they also produce the internal attributes of some of the elements in the arena, for example, the storage's resource level, number of resources each agent is carrying or whether the signal is active or not for any of the agents. To have a controllable setting, signals are produced by the environment at specific points in time (see below) and is provided as input to the agents. This creates a sandbox where we can evaluate how agents would learn and behave if signals are sent when different events happen. Finally, two learning agents with same action space interact with the environment.

Mining World: in this environment agents need to mine two type of resources and take them to a nearby storage. Mines have a probability of *0.01* of releasing a resource at each time step. Each agent has a limit of resources they can carry in its basket and every time step the level of each resource in the storage decreases linearly. Both agents get a reward of *0.1* every time step these resource levels are above zero, having the common goal of keeping the storage resources above zero. They also get an individual reward of *0.1* for each unloaded resource and, to be efficient, agents are expected to carry as many resources as possible. Additionally, if agents unload resource simultaneously they get a higher reward. For an example, see figure 2 left. If agents want to maximize the total reward, they will need to coordinate to unload resources at the same time, which becomes non-trivial due to the stochasticity of the mine mechanism to release resources and the limited field of vision of each agent. Here, an agent emits a signal when its basket is fully loaded.

Fort World: the dynamics of this environment are as follows: agents share the duty of protecting a fort, if this fort is destroyed both agents die. Additionally, they need to also gather food from four available points to keep the fort's food storage above zero and again, both agents die if food levels reach zero. Every time step the food level decreases on every time step but can be recovered. On the other hand, the fort's health level decreases when enemies attack it. These enemies can appear with a probability of *0.015* and can destroy the fort in twenty time steps, agents have to defeat their enemies by attacking them a fixed number of times. Here, if both agents collaborate, the enemies will die quicker getting more chances of survival but at the same time they need to keep gathering food. Note that agents do not need to return to the fort to collect resources, making the task of defending the fort more complicated. In this case, an agent sends a signal when it sees an enemy.

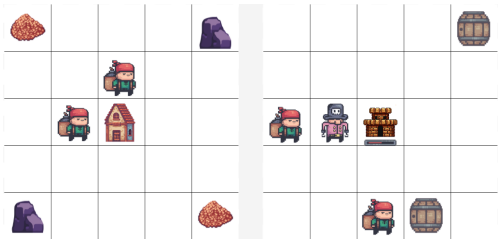

Figure 2: On the left, an example of the Mining World with two agents going to the storage to unload resources. On the right, the Fort World environment where one agent is gathering food and another agent is attacking an enemy assaulting the fort.

Measuring the Manipulation of Attention

Measuring the degree by which actions of an agent influence the attention of others is a non-trivial task. One way to do this would be to use the final collective reward since the more reward was collected the better the agents were collaborating. However, we argue that this methodology does not provide a clear and reliable measure, instead we suggest that to characterise the learning of this behavior in a reliable way, agents need to be able to answer the following question "Would have I acted in the same way if the other agent had not sent the signal?" i.e. use counterfactual knowledge (Pearl & Mackenzie,

2018). This means that if the behavior of an agent X changes due to agent Y's actions, agent Y can manipulate its attention. We note that in the case of the agents studied here, attention locus and location are in one-to-one correspondence. Following (Jaques et al., 2018), we reuse their intrinsic reward for causal influence to measure the degree to which an agent's signal s_l^i can manipulate the actions a^j of others:

$$p(a^j|o^j) \equiv \sum_{\forall k \neq l} p(a^j|s_k^i, o^j) p(s_k^i|o^i) \quad (1)$$

$$AM(a^j, s_l^i) \equiv D_{\mathrm{KL}}\left(p(a^j|s_l^i, o^j)||p(a^j|o^j)\right) \quad (2)$$

where $p(a^j|o^j)$ is the probability distribution over actions of agent j if the signal s^i would have taken any another value than s_l^i. D_{KL} is the Kullback-Leibler divergence (Kullback & Leibler, 1951) and $p(a^j|s_l^i, o^j)$ is the probability distribution of agent's j actions conditioned on seeing the observation o^j and agent i emitting the signal s_l^i. Note that the observation o^j is independent of the signal.

Results

We evaluate how relatively simple agents can manipulate the attention of others to achieve a common goal. This behavior is exemplified in figure 3 for the *Fort World* environment. Here, the signal is sent when one of the agents sees the enemy, making the other agent help defeat the enemy. In this case, the second agent would not have assisted the other agent without the emitted signal.

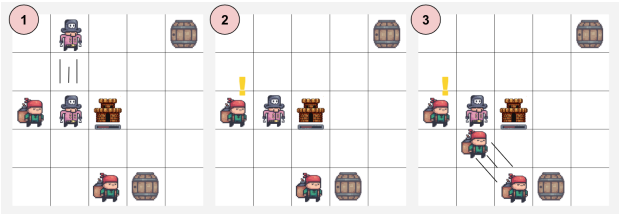

Figure 3: Example of attention manipulation in the *Fort World*. On the left, the enemy (pink character) comes to attack the fort. One agent (green and red) sees the enemy and tries to manipulate the attention of the other agent by sending a signal, middle. The second agent helps to avoid the assault by striking the enemy, right.

To demonstrate how attention manipulation can help in the *Mining World* environment, figure 4 exhibits the number of resources unloaded simultaneously by both agents, showing that in the case of using attention manipulation (orange) higher performance is achieved compared to not using this skill (blue). In figure 5, we show how our measurement for attention manipulation behaves in this same environment, capturing if agents are coordinating to unload resources simultaneously. For completeness, we also compute this metric for agents that do not use signals and, as expected, this measure gives low values when a signal is raised by another agent.

Figure 4: Number of resources unloaded simultaneously in the Mining World environment. Comparison between agents that can manipulate the attention of others (orange) and when they cannot (blue).

Surprisingly, in our experiments agents did not manipulate each other equally i.e. one agent would consistently manipulate the other agent more often, suggesting that one agent tends to follow what the other dictated.

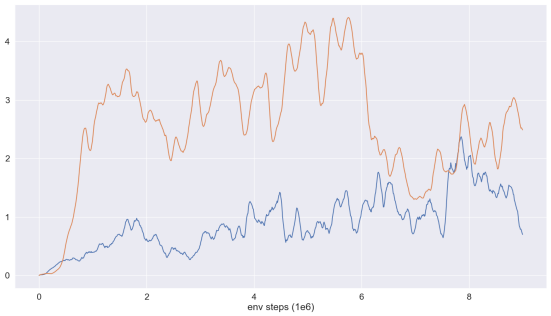

Figure 5: AM measurement applied in the Mining World to agents that can manipulate each others' attention (orange) and agents that cannot (blue). Each line combines the AM measurement for both agents.

In the case of the *For World* environment, we show in figure 6 how agents are able to defeat enemies (left) and collect food (middle) earlier in the training when they can manipulate each others' attention (orange) and that without (blue), agents take longer to find a good strategy to do both. The right side of the image shows the number of steps agents can keep alive in both settings and illustrates how they can stay alive longer during earlier stages of the training due to better coordination.

Discussion

Attention manipulation is an important prerequisite for complex coordinated behavior such as joint attention, which is thought to be essential for humans levels of cooperation and coordination. In this work, we have shown that RL agents

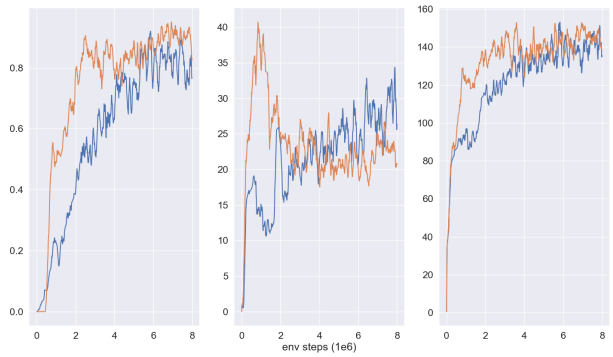

Figure 6: Comparison of agents with (orange) and without (blue) attention manipulation in the Fort World environment. Defeated enemies (left), collected food (middle) and number of steps alive (right) in average.

can use a similar behavior to attention manipulation when the environment is complex enough to provide a strong pressure for communication and cooperative behavior. We outlined some of the elements that make these environments complex enough for these pressures to be present and presented two environments with these elements. Additionally, we connected attention manipulation to causal influence and provided a way to measure how much agents can manipulate each others' attention using counterfactual and information theory measurements. Our results show that agents with simple architectures are able to use similar behavior to attention manipulation to accomplishing a common goal.

A limitation in our work to address is the fact that the signal is currently produced by the environment at a fixed point in time, this can be ineffective. We believe that by applying similar tools to our measure of attention manipulation will allow agents to send this signal when they consider to do so. Additionally, our environments are limited to two subtasks (e.g. collect food and defend fort). We will extend these environments to incorporate more sub-tasks, making the selection of which task an agent should attend to more complicated.

In conclusion, attention manipulation, as part of the joint attention framework, is a powerful tool used by humans to cooperate. As shown in this work, agents too can benefit from this tool to achieve complex tasks in a cooperative way. We believe that providing artificial agents with cognitive abilities like attention manipulation could open up important new research avenues.

References

Bolt, N. K., Poncelet, E. M., Schultz, B. G., & Loehr, J. D. (2016). Mutual coordination strengthens the sense of joint agency in cooperative joint action. *Consciousness and Cognition*, *46*, 173 - 187. doi: https://doi.org/10.1016/j.concog.2016.10.001

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, *63*(4), i. doi: 10.2307/1166214

Jaques, N., Lazaridou, A., Hughes, E., Gulcehre, C., Ortega, P. A., Strouse, D. J., ... De Freitas, N. (2018). *Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning* (Tech. Rep.).

Kaplan, F., & Hafner, V. V. (2004). *The Challenges of Joint Attention* (Tech. Rep.).

Kullback, S., & Leibler, R. A. (1951, 03). On information and sufficiency. *Ann. Math. Statist.*, *22*(1), 79–86. doi: 10.1214/aoms/1177729694

Moore, C., & Dunham, P. J. (1995). *Joint attention: Its origins and role in development*. Lawrence Erlbaum.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect* (1st ed.). New York, NY, USA: Basic Books, Inc.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms. doi: 10.1007/s00038-010-0125-8

Tomasello, M. (2019). Becoming Human: a Theory of Ontogeny.