Tailoring Exploration and Exploitation in Multi-Agent Systems with Short-Term Memory and Limited Social Interaction

Hian Lee Kwa[1,2], Jabez Leong Kit[1] and Roland Bouffanais[3]

[1]Singapore University of Technology and Design, Singapore
[2]Thales Solutions Asia, Singapore
[3]University of Ottawa, Canada
hianlee_kwa@mymail.sutd.edu.sg

A common challenge faced by all decentralized multi-agent systems (MAS) is the exploration-exploitation dilemma. This stems from the fact that gathering new information about the environment (i.e., exploration) and making use of currently known information (i.e., exploitation) tend to be mutually exclusive activities. While heavily biasing an MAS towards exploration would allow large amounts of data to be gathered, it would be prevented from fully benefiting from this information. Conversely, over-exploitation may yield fast convergence times but can also result in agents being trapped in local optima or being unable to adapt to dynamic environments. Therefore, maximizing the performance of a system, especially when operating in dynamic environments, requires some sort of regulation of the amount of exploration and exploitation carried out by an MAS. In this work, we outline how the strategy developed in our work (Kwa et al., 2021) allows for the regulation and control of a swarm's exploration and exploitation dynamics (EED) within the context of tracking fast-moving evasive and non-evasive targets.

Previously, Esterle and Lewis (2020) have shown that increasing the level of inter-agent communications improved the overall tracking performance of the system while tracking slow-moving targets. However, it has been demonstrated that this is not the case when tracking fast-moving non-evasive targets (Kwa et al., 2020). This is because higher levels of inter-agent connectivity leads to over-exploitation of a target's positional information, causing agents to lose track of the target once they have been outrun. Similarly, low levels of connectivity result in too much exploration and insufficient exploitation, preventing agents from effectively tracking the target. As such, an optimal level of connectivity exists at which the level of exploration and exploitation carried out by the MAS was relatively balanced, maximizing its tracking performance. Similar optimal levels of connectivity maximizing system performance were found in other scenarios such as in obtaining a dynamic consensus (Mateo et al., 2019) and in a collaborative stick pulling task (Hamann, 2018).

In Kwa et al. (2020), a tracking strategy was proposed based on a memory-less Particle Swarm Optimization (PSO) algorithm, enabling the tracking of a fast-moving non-evasive target. However, this strategy does not allow the tracking of fast-moving evasive targets due to insufficient exploitation. In this work, we demonstrate the pivotal role played by short-term memory when tracking fast-moving evasive targets that can move faster than the individual constituent agents of the tracking MAS. Its introduction also gives another parameter that can tune an MAS' EED in conjunction with the level of inter-agent connectivity. This tuning allows the system to adapt its collective dynamics to maximize its performance while tracking different target numbers, speeds, and movement profiles.

The proposed strategy essentially consists of two behaviors: (1) promotion of agent aggregation around a point of attraction (exploitation), and (2) an adaptive inter-agent repulsion behavior (exploration). These behaviors generated two velocity vectors at each time-step that were combined to give a final agent velocity vector:

$$\mathbf{v}_i[t] = \mathbf{v}_{i,\text{att}}[t] + \mathbf{v}_{i,\text{rep}}[t], \tag{1}$$

where $\mathbf{v}_{i,\text{att}}[t]$ and and $\mathbf{v}_{i,\text{rep}}[t]$ are the velocity vectors generated by the attractive component and the repulsion component respectively. Selecting the degree, k, of the inter-connecting k-nearest neighbor communications network controlled the amount of social interaction between the swarming agents, and hence the overall EED of the swarm.

The main difference between the algorithm presented here and the one in Kwa et al. (2020) is the addition of a short-term memory to the agent aggregation component, thus enabling the tracking of evasive targets. This attractive component generates an attractive velocity vector:

$$\mathbf{v}_{i,\text{att}}[t+1] = \omega \mathbf{v}_i[t] + cr\big(\mathbf{p}[t+1] - \mathbf{x}_i[t+1]\big), \tag{2}$$

where $\mathbf{p}$ is the most recent target position as observed by an agent and its neighbors, $\mathbf{x}_i$ is the position of agent i, ω is the velocity inertial weight, c is the social weight, and r is a number randomly drawn from the unit interval.

In the pursuit of an fast-moving evasive target, the use of agent-based memory gives the swarm a longer lasting point of attraction. This increases the amount of exploitation carried out by the MAS, allowing it to close in on a target even though agents are unable to detect the presence of the target. As such, each agent is given a memory, M, with a duration of t_{mem}. Should the amount of time elapsed since $\mathbf{p}$ was last updated exceed t_{mem}, $\mathbf{p}$ will be cleared by setting $\mathbf{p} = \mathbf{x}_i[t]$, removing the point of attraction from consideration when calculating the final velocity vector. Doing so causes the agents to move away from each other, expanding their formation, thereby carrying out area exploration. Therefore, by lengthening the duration of t_{mem}, an MAS will carry out a higher level of exploitation while shortening the duration of t_{mem} will shift the EED balance in favor or exploration.

To study the performance of the system, we simulated an MAS with 50 agents tracking a single evasive and non-evasive target over a period of 100,000 time-steps and calculated the percentage of time that the target was tracked (performance measure Ξ). An *Engagement Ratio*, the proportion of time all agents spend actively trying to move towards a target, was also calculated to quantify the MAS' EED.

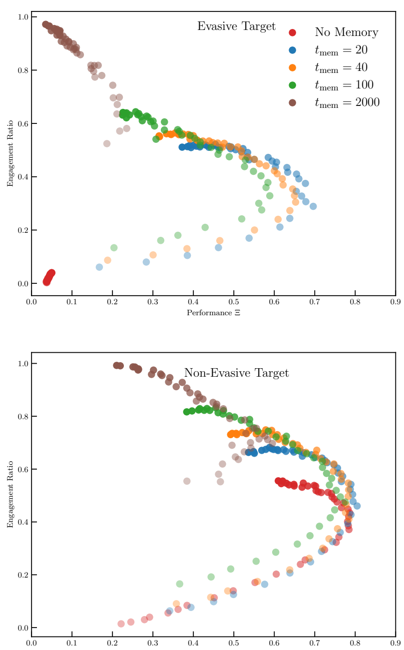

Figure 1: Engagement-Tracking plots of a swarm with different t_{mem} and k tracking an evasive (top) and a non-evasive (bottom) target. Darker shaded points indicate swarms using networks with higher values of k.

As seen in Fig. 1 not only is there an optimal level of connectivity at which an MAS' tracking performance is maximized, there is also an optimal level of engagement at where this maximum occurs. The optimum engagement is also higher when tracking non-evasive targets compared to when tracking evasive targets. This is because an evasive target makes its movements to avoid contact with pursuing agents, resulting in more exploration demanded from the MAS to track the target, which is associated with lower engagement ratios. Therefore, reducing the MAS' level of connectivity reduces the system's engagement, leading to better tracking performances when tracking evasive targets.

The figures also demonstrate the crucial role of short-term memory in facilitating the tracking of an evasive target. Without its presence, it can be seen that the MAS was unable to track the evasive target. This is because the introduction of short-term memory generates a persistent point of attraction based on a target's last known position, giving the agents the ability to aggregate at that point even though the target may have moved away. Without memory, while the pursuing agents may periodically encounter the target, the agents repel each other and expand until the system reaches a static equilibrium position and is unable to close in on the target because of the target's evasive maneuvers. Similarly, the plots also illustrate the detrimental nature of using long-term memory. The use of longer memory lengths, which is associated with higher levels of exploitation, causes the MAS to over-exploit outdated target information, resulting in agents aggregating at a location where the target is no longer present, hence leading to lower tracking performance.

Future work in this field will involve developing and tailoring agent behaviors and communication topologies for use in varying swarm densities. In addition, the use of heterogeneous MAS can also be studied to take advantage of different agent behaviors and physical capabilities.

Acknowledgements

This research is funded by Thales Solutions Asia Pte Ltd, under the Singapore Economic Development Board IPP.

References

Esterle, L. and Lewis, P. R. (2020). Distributed Autonomy and Trade-Offs in Online Multiobject k-Coverage. *Computational Intelligence*, 36(2):720–742.

Hamann, H. (2018). Superlinear scalability in parallel computing and multi-robot systems: Shared resources, collaboration, and network topology. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10793 LNCS:31–42.

Kwa, H. L., Kit, J. L., and Bouffanais, R. (2020). Optimal Swarm Strategy for Dynamic Target Search and Tracking. In *Proc. 19th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2020)*, pages 672–680, Auckland, New Zealand.

Kwa, H. L., Kit, J. L., and Bouffanais, R. (2021). Tracking Multiple Fast Targets With Swarms: Interplay Between Social Interaction and Agent Memory. In *ALIFE 2021: The 2021 Conference on Artificial Life*, Prague, Czech Republic.

Mateo, D., Horsevad, N., Hassani, V., Chamanbaz, M., and Bouffanais, R. (2019). Optimal Network Topology for Responsive Collective Behavior. *Sci. Adv.*, 5:eaau0999.