FRIteza

Swarming behaviour in predator-prey model

Ariana Kržan, Tina Brdnik, and Vito Levstik

Collective behaviour course research seminar report

January 10, 2025

Iztok Lebar Bajec | associate professor | mentor

Collective animal behaviour, especially swarming in predator-prey dynamics, offers insights into survival strategies that emerge under evolutionary pressures. This report outlines the initial objectives and foundational concepts for simulating predator-prey. Inspired by previous work, we examined how survival pressures can drive emergent group behaviours in prey through reinforcement learning. Our primary objective was to recreate a reinforcement learning-based model where predator-prey interactions lead to swarming and evasion behaviours. Unfortunately, we have not been able to observe any swarming behaviour. We suspect that the rewards used in the model lead to suboptimal results. With different learning parameters we might be able to achieve desired results, but so far we have not been able to do so. The model was then extended to include environmental obstacles and an additional species, to provide framework for future work on interspecies interactions and new survival strategies.

Simulation | swarming behaviour | predator | prey

The sudden emergence of swarming behaviours in animals is one of the most striking examples of collective animal behaviour. These behaviours have been extensively studied for their implications for the evolution of cooperation, social cognition and predator–prey dynamics [1]. Swarming, which appears in many different species like starlings, herrings, and locusts, has been linked to several benefits including enhanced foraging efficiency, improved mating success, and distributed problem-solving abilities. Furthermore, they are hypothesized to help with improving group vigilance, reducing the chance of being encountered by predators, diluting an individual's risk of being attacked, enabling an active defence against predators and reducing predator attack efficiency by confusing the predator [2].

In this project we took the inspiration from the work of Li et al. (2023) [2] to explore how simple survival pressures can drive the emergence of swarming behaviour. The first goal was to create a realistic simulation where both prey and predators learn to adapt through reinforcement learning based on their drive to survive. Unfortunately, we have not yet been able to recreate the same results as in [2] or at least observe any swarming behaviour.

Nevertheless, we then extend our work by providing framework for the introduction of new environmental obstacles and new species with the desire to observe how interspecies interactions lead to new survival strategies [3].

Methods

Our proposed methodology aimed to simulate swarming behaviours in a predator-prey environment using reinforcement learning (RL). We defined and tested a RL-based model where agents, such as prey and predators interacted within a two-dimensional space. The goal was to observe collective behaviours like swarming, evasion, and strategic movement.

We implemented our own environment setup, which we wanted to compare with the model from Li et al. (2023) [2].

Environment Setup. The simulation took place in a 2D environment with open space, meaning that agents reappeared on the opposite side when they crossed the boundary. Such setup with periodic boundaries serves as an approximation of an infinite space, where agents are allowed to move freely without encountering physical borders. Later on, we introduced a new species to the environment and placed random obstacles, which simulated more realistic and complex space that challenged the agents to adapt their movement and coordination.

Agent Dynamics. Agents in our simulations were subject to both active and passive forces.

Active forces are self-generated by agents to drive their movement. These forces consist of two components:

- **Forward Propulsion**: Drives the agent in the direction of its heading. This force is represented as a_F.

This project investigates the emergence of swarming behaviors in predator-prey dynamics using reinforcement learning. Despite challenges in replicating swarming behaviors, the work establishes a foundation for exploring the impact of environmental obstacles and additional species on survival strategies. By expanding the model to include interspecies interactions and complex environments, this study provides a framework for future research into collective behavior and adaptive strategies in multiagent systems.

Simulation | swarming behaviour | predator | prey

- **Rotational Force**: Allows the agent to rotate its heading within a threshold value. This force is denoted as a_R, where a_R controls the angular velocity.

Passive forces act on agents due to interactions with the environment and other agents. These include:

- **Dragging Force**: Acts opposite to the agent's velocity, simulating frictional effects. It is proportional to the magnitude of the velocity $\mathbf{v}$.
- **Elastic Forces Between Agents**: When agents are in contact, elastic forces prevent overlap and simulate collision dynamics. These forces follow Hooke's law and are represented as $\mathbf{f}_a$.

The RL framework aimed to optimize the agents' use of active forces a_F and a_R to maximize their survival behaviors.

Agent Types and Behaviour.

- **Prey**: Prey aims to survive by avoiding predators and moving as a group.
- **Predators**: Predators are designed to pursue and catch prey.
- **Super Predators**: We introduced a third type of agent called super predators. The aim of super predators is to catch predators, while not caring about the prey. This would simulate a real world food chain, where predators would have to choose between catching prey or avoiding super predators.

Reinforcement Learning Framework. At first we wanted to use the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm [2, 4], but we had a hard time writing it from scratch. Hence, we later decided to use Proximal Policy Optimization (PPO) algorithm [5].

We followed the same reward structure as in Li et al. (2023) [2], where the reward for prey is set to -1 if it is caught by a predator and the predator is awarded $+1$ if it catches a prey. As in the original article we also added small penalty of $-0.01|a_F| - 0.1|a_R|$ which mimics the cost of movement. When adding super predators, we followed the same reward structure for the interaction between predators and super predators. We also added obstacles and added a toggable negative reward if any agent touched it, so we could observe it's effects.

Agents were trained through episodic simulations, allowing them to learn and adapt from each episode's interactions. We varied parameters to try and get the best results, but so far we have not been successful.

Proposed Methodology for Verification. To verify the behavior of our model, we adopted the methodology described in Li et al. (2023) [2], utilizing two key metrics: the Degree of Alignment (DoA) and the Degree of Separation (DoS).

- **Degree of Sparsity (DoS):** This metric measures the spatial aggregation of agents, capturing how densely the agents cluster together. It is defined as:

$$\text{DoS} = \frac{1}{TND} \sum_{t=1}^{T} \sum_{j=1}^{N} \|\mathbf{x}_j(t) - \mathbf{x}_k(t)\|$$

  where: $\mathbf{x}_j(t)$ is the position of the j-th agent at time step t, $\mathbf{x}_k(t)$ is the position of the nearest neighbor $k = \arg\min_k \|\mathbf{x}_j(t) - \mathbf{x}_k(t)\|$, T is the episode length, N is the total number of agents, and D is the maximum possible distance between two agents in the environment.

  A smaller DoS value indicates denser clustering, while a value of 0 represents all agents aggregating at a single point [2].

- **Degree of Alignment (DoA):** This metric quantifies the alignment of the agents' headings, assessing how consistently agents move in the same direction. It is defined as:

$$\text{DoA} = \frac{1}{2TN} \sum_{t=1}^{T} \sum_{j=1}^{N} \|\mathbf{h}_j(t) + \mathbf{h}_k(t)\|$$

  where: $\mathbf{h}_j(t)$ is the heading of the j-th agent at time step t, $\mathbf{h}_k(t)$ is the heading of the nearest neighbor of agent j (the same nearest neighbor as in the DoS calculation), T is the episode length, and N is the total number of agents.

  Higher DoA values indicate stronger alignment in agent movement. It is important to note that the DoA measures local alignment between neighboring agents rather than the mean heading of the entire group, making it more suitable for detecting relative alignment within swarms [2].

Results

In the initial phase of our project, we implemented a basic model where we created our environment with periodic borders and successfully populated it with agents which followed a reward system following the article. At this stage we only implemented active forces with fixed values. Our results looked promising.

Next we improved the model by first adding passive forces. We also fine-tuned some parameters such as the agents' speed, size and passive forces. The passive forces included dragging forces and elastic forces between contacting agents. Then we initiated active forces as random instead of fixed valued. That was crucial for our next step, since we were going to build our reinforcement model based on them.

Lastly we added a RL component. The RL component was built to fine-tune the active forces a_f and a_f to enable the agents to move efficiently and maximize their rewards. We tried implementing a MADDPG algorithm like in the article. However after training our model numerous times over 1000 episodes, the results were not great. Instead of the agents moving to maximize their reward, they ended up moving in circles.

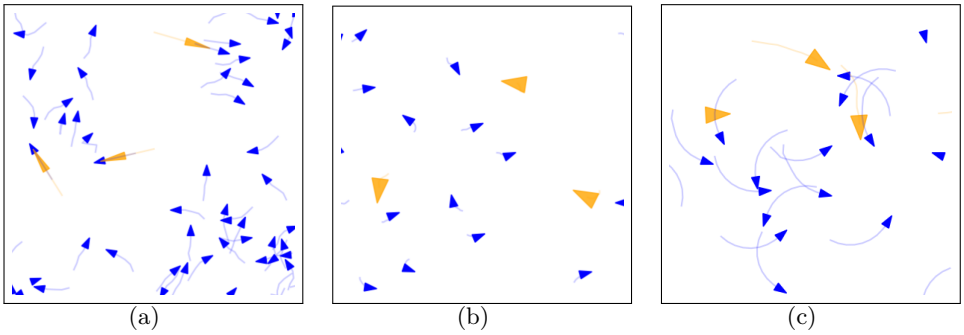

Figure 1. (a) Our model with no RL component, following the reward system and fixed active forces. (b) Our model with no RL component, with added fixed passive forces and random initialization of active forces (c) Our model with RL, after training.

Due to our unsatisfactory results, we decided to adopt the RL model from the Li et al. article [2]. Using their code, we were able to successfully run simulations of their model in our environment that can be seen in figure 2. This provided a functioning baseline for comparison and further experimentation. Our results demonstrate significant progress in achieving swarming behaviours. Specifically, we achieved an impressive Degree of Separation (DoS) value of 0.12. The prey agents were able to form tight clusters while escaping the predators. However, the Degree of Alignment (DoA) results remained suboptimal, with values consistently staying at around 0.3. This indicates that while agents were able to group closely, they struggled to align their velocities and directions cohesively. Even after consulting with the original authors we didn't get better results.

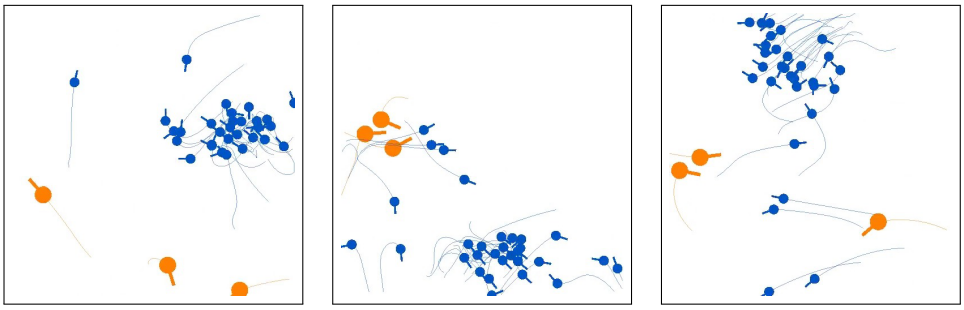

Figure 2. Visualization of simulation results showing agent behavior. The prey demonstrate successful swarming behavior, as indicated by the tight clustering of prey, resulting in a low Degree of Separation. However, the also show poor alignment, leading to low DoA numbers.

We tried modifying their code to add obstacles and the third species but we failed to do so to the complex nature of their code. That is why we tried fixing our model one last time. This time we used the PPO algorithm. Using the same reward structure we observed agents becoming stationary, but when we decreased the penalty for movement, the agents began moving normally. Still we were not able to observe any swarming behaviour. Visually, this model is much better than the previous one we

have implemented, as the agents do not move in circles, but the results are still not satisfactory. The fact that the agents become stationary with the original reward structure is somewhat expected, as agents simply learn to avoid moving to avoid penalties. This makes us thinking that using different training parameters might lead to results similar to the ones in the article [2]. Unfortunately, we have not been able to find the right parameters that would lead to the desired results.

On this model we also added obstacles and the third species. Learning in such environment again did not produce any swarming behaviour. The results can be seen in figure 3.

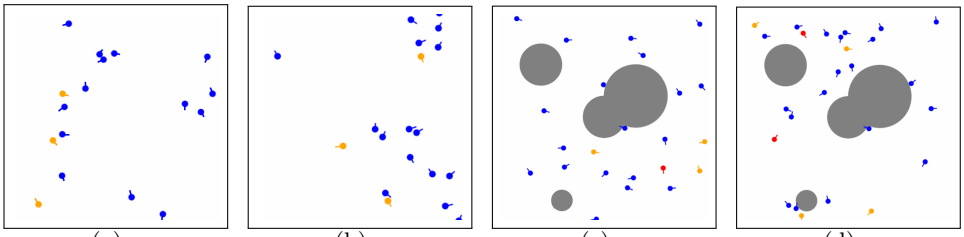

Figure 3. (a) Final model before training. (b) Final model after training. (c) Final model before training with obstacles and superior predator. (d) Final model after training with obstacles and superior predator.

Discussion

In this stage of the project, we successfully ran the code from the main article but encountered challenges with the agents' behavior. The DoA and DoS metrics remain highly variable, indicating that swarming behaviors have not yet emerged. These results suggest the need for further tuning of the reward structure and additional training.

Despite the setbacks, we are motivated to continue refining the model and addressing these issues. Our next steps are to improve the code to produce results similar to the article, introduce obstacles to the environment, and add a species, which we have yet to finalize.

CONTRIBUTIONS. AK worked on the first two models and writing agent dynamics and results, TB worked on graphs, methods, results and discussion, VL worked on original model, implementing PPO model and methods.

1. Olson RS, Hintze A, Dyer FC, Knoester DB, Adami C (2013) Predator confusion is sufficient to evolve swarming behaviour. *Journal of The Royal Society Interface* 10(85):20130305.
2. Li J, Li L, Zhao S (2023) Predator–prey survival pressure is sufficient to evolve swarming behaviors. *New Journal of Physics* 25(9):092001.
3. Sapkota N, Bhatta R, Dabney P, Xie Z (2020) Hunting cooperation in the middle predator in three species food chain model. *arXiv preprint arXiv:2006.16525*.
4. Li S et al. (2019) Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient in *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33, pp. 4213–4220.
5. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O (2017) Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.