Evolving Dynamic Collective Behaviors by Minimizing Surprise

Tanja Katharina Kaiser[1,2], Christopher Kluth[3], and Heiko Hamann[4]

[1]Center for Interdisciplinary Digital Sciences, TU Dresden, Germany
[2]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, Germany
[3]Institute of Computer Engineering, University of Lübeck, Germany
[4]Department of Computer and Information Science, University of Konstanz, Germany

Abstract

Our minimize surprise method evolves swarm robot controllers using a task-independent reward for prediction accuracy. Since no specific task is rewarded during optimization, various collective behaviors can emerge, as has also been shown in previous work. But so far, all generated behaviors were static or repetitive allowing for easy sensor predictions due to mostly constant sensor input. Our goal is to generate more dynamic behaviors that vary behavior based on changes in sensor input. We modify environment and agent capabilities, and extend the minimize surprise reward with additional components rewarding homing or curiosity. In preliminary experiments, we were able to generate first dynamic behaviors through our modifications, providing a promising basis for future work.

Introduction

Evolutionary algorithms are a promising approach to automatically generate collective behaviors for robot swarms (Trianni, 2008). Our minimize surprise method (Hamann, 2014) rewards high prediction accuracy instead of a specific task. Consequently, various swarm behaviors can emerge in the optimization process. So far, the swarm had full controllability (Jung et al., 2011) over the environment as arenas contained only the swarm or the swarm and passive building material (Kaiser and Hamann, 2022). As all interactions resulted from the agent behavior itself, repetitive or static behaviors emerged in these experiments that are relatively easy to predict due to constant sensor input. In this article, we aim for the evolution of dynamic behaviors that can be used in more complex scenarios by minimizing surprise. We study the impact of modifications to the environment and agent capabilities as well as extensions to the reward on the emergence of behaviors. Our inspiration comes from studies focusing on adaptation to environmental influences and survivability. Similar to Miras et al. (2020) and Berseth et al. (2020), we investigate the effect of environmental conditions by testing both a static environment (except for the behavior of the swarm itself) and an environment that is dynamic independent from agent behavior. In robot ecology, Egerstedt et al. (2018) focus on the survivability of agents to enable long-duration autonomy. We adopt their idea of introducing ecological constraints. We limit the agent's simulated battery and require recharging (Lowe et al., 2010), which corresponds to an animal's need to find nutrition to survive. We study two options of extending our standard minimize surprise reward for prediction accuracy: (a) with an additional task-specific reward for homing and (b) with a task-independent reward for curiosity (Schmidhuber, 1991). First dynamic behaviors emerged in our scenarios providing a promising basis for future work.

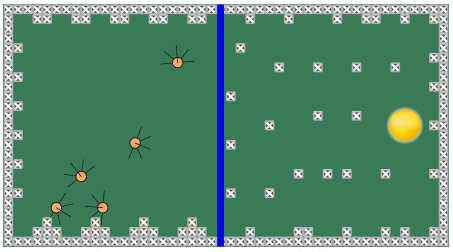

Figure 1: Arena with agents (orange circles, black lines indicate proximity sensors), obstacles (gray squares), and light source (yellow circle). The blue line marks our division of the arena into a safe zone and a risk zone.

Approach

We use a swarm of 5 simulated agents in a 2D arena bounded by walls, see Fig. 1. The arena is divided into two zones introducing spatial variability: The safe zone has blocks along the walls that are placed in unique patterns per compass direction. The risk zone contains 34 randomly placed obstacles and a light source that serves as a charging station for the agent batteries. Light intensity $I \in [0, 1]$ decreases with distance to the light source. Agents can move forward or rotate and have a battery initially lasting for 2 000 time steps. Each agent has 8 sensors: 5 proximity sensors, 1 light intensity sensor, 1 compass, and 1 battery sensor. Collisions are avoided by a simple hardware protection layer enforcing rotations on spot if an agent's forward movement

Table 1: Tested settings

reward	RoSD	ES	CE	mean fitness $\pm$ STD
F_{MSH}	-	-	-	0.76 ± 0.04
F_{MSC}	-	-	-	0.82 ± 0.03
	$\times$	-	-	0.71 ± 0.07
	-	$\times$	-	0.77 ± 0.08
	$\times$	$\times$	-	0.79 ± 0.02
	-	$\times$	$\times$	0.82 ± 0.02

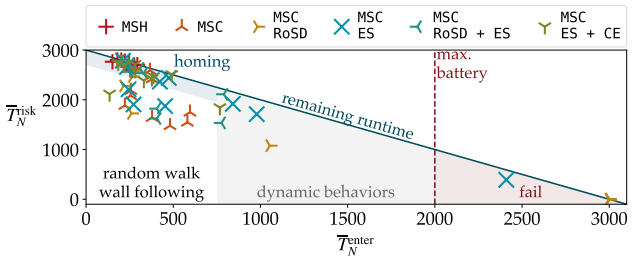

Figure 2: Behavior space for the different tested settings based on times agents spent in the risk zone $\overline{T}_N^{\text{risk}}$ and time $\overline{T}_N^{\text{enter}}$ until agents first enter the risk zone.

would result in a collision. Following our minimize surprise method (Hamann, 2014), each agent is equipped with an actor-predictor pair of artificial neural networks. The actor outputs a linear and angular velocity while the predictor determines predictions for the agent's proximity sensors and its light sensor. Inputs to both networks are all sensor values and the current or next linear and angular velocities.

We study the effect of 3 different agent and environment modifications, and combinations thereof (see Tab. 1): *Energy Sharing* (ES) enables agents to equally share their battery levels when in contact. *Risk of Sudden Discharge* (RoSD) increases the hazard in the risk zone by introducing a probability that an agent's battery suddenly discharges completely, rendering the agent useless. *Changing Environment* (CE) lets the positions of obstacles and light source in the risk zone change in equal time intervals.

In addition, we test 2 different rewards that rely on prediction accuracy (i.e., they minimize the difference between the actual sensor value and the predicted sensor value). Reward F_{MSC} rewards curiosity in addition to prediction accuracy by punishing constant sensor input. Agents receive a reward of either the prediction accuracy or, if it is lower, of $1 - \rho$ per time step, whereby $\rho \in [0, 1]$ linearly increases between 150 and 350 time steps with constant sensor values. Reward F_{MSH} rewards homing in addition to prediction accuracy. It is a weighted sum of prediction accuracy and detected light intensity as a measure of the agent's distance to the light source where the weighting depends on the battery level. Proximity to the light source is rewarded the higher, the lower the battery level, increasing the chance that agents recharge their batteries on time.

We do 10 independent evolutionary runs per setting using a simple evolutionary algorithm with population size 50, elitism of one, a mutation rate of 0.4, and no crossover. Each individual is evaluated once for $T = 3\,000$ time steps and we terminate evolution after 150 generations.

Results

In all tested settings, we find a mean best fitness between 0.71 and 0.82 indicating the successful optimization of the actor-predictor pairs, see Tab. 1. We analyze the effect of the different rewards and modifications on the emergent behaviors, see Fig. 2. All best-evolved individuals in the setting using reward F_{MSH} lead to homing which is most likely caused by the reward's task-specific component for reaching the light source. Thus, we do not investigate this reward further. In all settings using reward F_{MSC}, we find a majority of static or repetitive collective behaviors. Most frequent are homing, random walk, and wall-following behaviors. For all of these behaviors, agents enter the risk zone quickly ($\overline{T}_N^{\text{enter}} < 750$). For homing, agents then stay the rest of the evaluation in the risk zone ($T - \overline{T}_N^{\text{enter}} \approx \overline{T}_N^{\text{risk}}$). For random walk and wall following, agents leave the risk zone again ($T - \overline{T}_N^{\text{enter}} \gg \overline{T}_N^{\text{risk}}$). In addition, we find 5 best-evolved individuals that still fail to recharge their batteries in time ($\overline{T}_N^{\text{enter}} > 2000$). A few best-evolved individuals lead to dynamic behaviors. Agents enter the risk zone late enough ($\overline{T}_N^{\text{enter}} > 750$) s. t. a single recharge is sufficient to survive the full run. They switch the location where they exhibit a random walk or circling behavior based on battery level. All of these behaviors emerged in settings with agent or environment modifications. Consequently, external pressures from the environment and ecological constraints affecting the agent influence the evolution of dynamic behaviors.

Conclusion

While we only find a few dynamic behaviors in our experiments, the results highlight the importance of ecological constraints and dynamics in the environment to push evolution toward the emergence of dynamic behaviors when using a task-independent reward. We only study small modifications of environment and agents that have only a limited effect on emergent behaviors. In future work, we will study disruptive forces in the environment by introducing, e.g., seasons or independently acting and potentially malicious agents.

Acknowledgements

This work was partially supported by the German Federal Ministry of Education and Research (BMBF, SCADS22B) and the Saxon State Ministry for Science, Culture and Tourism (SMWK) by funding the competence center for Big Data and AI "ScaDS.AI Dresden/Leipzig".

References

Berseth, G., Geng, D., Devin, C., Rhinehart, N., Finn, C., Jayaraman, D., and Levine, S. (2020). SMiRL: Surprise minimizing reinforcement learning in dynamic environments. arXiv.

Egerstedt, M., Pauli, J. N., Notomista, G., and Hutchinson, S. (2018). Robot ecology: Constraint-based control design for long duration autonomy. *Annual Reviews in Control*, 46:1–7.

Hamann, H. (2014). Evolution of collective behaviors by minimizing surprise. In Sayama, H., Rieffel, J., Risi, S., Doursat, R., and Lipson, H., editors, *14th International Conference on the Synthesis and Simulation of Living Systems (ALIFE 2014)*, pages 344–351. MIT Press.

Jung, T., Polani, D., and Stone, P. (2011). Empowerment for continuous agent—environment systems. *Adaptive Behavior*, 19(1):16–39.

Kaiser, T. K. and Hamann, H. (2022). Innate motivation for robot swarms by minimizing surprise: From simple simulations to real-world experiments. *IEEE Transactions on Robotics*, 38(6):3582–3601.

Lowe, R., Montebelli, A., Ieropoulos, I., Greenman, J., Melhuish, C., and Ziemke, T. (2010). Grounding motivation in energy autonomy: A study of artificial metabolism constrained robot dynamics. pages 725–732. Massachusetts Institute of Technology Press (MIT Press).

Miras, K., Ferrante, E., and Eiben, A. E. (2020). Environmental influences on evolvable robots. *PloS one*, 15(5):e0233848.

Schmidhuber, J. (1991). A possibility for implementing curiosity and boredom in model-building neural controllers. In *Proceedings of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, pages 222–227.

Trianni, V. (2008). *Evolutionary Swarm Robotics - Evolving Self-Organising Behaviours in Groups of Autonomous Robots*, volume 108 of *Studies in Computational Intelligence*. Springer, Berlin, Germany.