Information-Efficient Vicsek Flocking Using Deep Reinforcement Learning

Jongyun Kim[1], Hae-In Lee[1], Hyo-Sang Shin[1 2], and Antonios Tsoudos[1]

[1] Cranfield University, College Rd, Cranfield, Bedford, MK43 0AL, UK
{jongyun.kim,haein.lee,h.shin,a.tsourdos}@cranfield.ac.uk,
[2] KAIST, 193, Munji-ro, Yuseong-gu, Daejeon, 34021, Republic of Korea
hyosangshin@kaist.ac.kr

Abstract. This paper presents an approach to optimizing information-efficient flocking in a multi-agent system using deep reinforcement learning. Inspired by biological observations, our method focuses on enhancing collective motion by selectively reducing the neighbor information for the Vicsek model processed by each agent. The proposed learning-based framework autonomously learns the optimal neighbor selection probability, effectively reducing information overhead without compromising the flocking behavior. Numerical simulations demonstrate that our approach achieves faster convergence and superior alignment with significantly reduced information utilization compared to the traditional method.

Keywords: collective behavior, intelligent control, neighbor selection

1 Introduction

Collective behaviors in nature often involve individuals within a group making decisions based on neighbor information, leading to coordinated global behavior. However, not all individuals actively participate in this coordination, which can slow down the emergence of collective behavior. For instance, many ants in a colony are inactive [1], and certain giant honeybees in South Asia show refractory periods of inactivity during coordinated shimmering [2].

Despite these apparent inefficiencies, these behaviors might be advantageous, as they are common in nature. Many species utilize selective reinforcement from specific neighbors rather than all available information. For example, fish adjust their social networks to prioritize certain neighbors, enhancing information efficiency [3,4]. Similarly, in human social networks, people are more influenced by diverse social groups [5]. Studies on starlings show that reducing the amount of neighbor information can improve group alignment [6].

Applying these principles to robotics, various studies have explored neighbor selection in flocking models. Lu et al. [7,8] enhanced the Vicsek model by weighting neighbor information based on connections or distance, improving alignment and convergence. Other researchers, such as Martin and Zhang, have investigated topological interactions to improve synchronization [9,10].

Despite the extensive research, the optimal neighbor selection strategy remains unclear, often requiring expert knowledge and assumptions. To address this, we propose a data-driven method using deep reinforcement learning (DRL) for neighbor selection in flocking. This approach autonomously learns optimal information selection, reducing the need for manual strategy design and assumptions, and is validated through numerical simulations.

2 Information-Efficient Flocking

2.1 Vicsek Flocking Model and Scenarios

The Vicsek Flocking Model [11] describes the collective motion of self-propelled agents, each moving with a constant speed v and updating its velocity based on the average velocity of its neighbors within a communication radius r. The positions $\mathbf{x}_i(t)$ and velocities $\mathbf{v}_i(t)$ of agents evolve over discrete time steps t as follows: $\mathbf{v}_i(t+1) = \frac{1}{n_i} \sum_{j \in N_i} \mathbf{v}_j + \eta\xi_i(t)$ and $\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1) \cdot \Delta t$, where N_i represents the neighbors of agent i within distance r, $n_i = |N_i|$, and $\eta\xi_i(t)$ accounts for random noise. In this model, N agents interact within a periodic boundary condition, allowing agents near the edges to interact with those on the opposite side. The degree of alignment among agents is quantified by the order parameter $\phi(t)$, defined as: $\phi(t) = \frac{1}{Nv} \left\| \sum_{i=1}^{N} \mathbf{v}_i(t) \right\|_2$, where $\phi(t) = 1$ indicates perfect alignment and $\phi(t) \approx 0$ indicates a disordered state. This parameter serves as a crucial metric for evaluating flocking behavior.

2.2 Deep Reinforcement Learning-Based Selector

The core principle of the neighbor selection strategy is to selectively disregard less relevant neighbor information, enhancing efficient flocking convergence while keeping the underlying model unchanged. Given the complexity of manually designing such a policy, the proposed approach leverages Deep Reinforcement Learning (DRL) to automate this decision-making process. In this approach, each agent's policy network receives information about neighboring agents, such as their states, and calculates a selection probability for each neighbor. A subset of this information is then stochastically selected and passed to the flocking algorithm, which transforms the reduced information into control inputs, leading to faster collective behavior.

This DRL-based method is framed as a Markov decision process. The agent's state s_i represents the message to be delivered to other agents, simplified as the dynamics state $[\mathbf{x}_i, \mathbf{v}_i]^T$. The action a_i consists of the selection probabilities for each neighboring agent, determined by the policy network. The state transitions follow the Vicsek model, and the reward function $R(s, a) = \phi(t)$, the order parameter, incentivizes rapid flocking formation.

The policy network uses an encoder-decoder architecture with attention mechanisms [12]. The encoder processes the input states through a linear embedding layer, followed by self-attention and feedforward layers. The output is a mean embedding representing the flock's contextual state, which is then concatenated with each encoder output. The decoder includes a cross-attention layer, where the query comes from the flock context, and the keys/values come from the encoder outputs. The resulting scores are processed to generate selection probabilities through a softmax function. The policy is updated using the proximal policy optimization [13]. A separate value network, with a structure similar to the policy network, estimates the value. Despite the complexity, all agents share a single neural network with common parameters, ensuring computational efficiency and consistency in a cooperative task where agents have identical observation and action spaces.

3 Numerical Simulation

3.1 Simulation Settings

A numerical simulation was performed to validate the proposed approach. Twenty agents were uniformly distributed within a 100×100 m area, each with a random heading. The simulation used a 0.1 s time step, with episodes lasting 30 s. Agents had a communication radius of 10 m and moved at a constant speed of 5 m/s. Network configurations including the architecture and the hyperparameters can be found on GitHub: https://github.com/JongYun-Kim/lazy-vicsek.git.

In the simulation, three evaluation metrics were used: episode reward, final order parameter, and information utilization. The episode reward R_{ep} is the sum of the order parameters over time, representing overall alignment capability. The average episode reward $\overline{R}_{\mathrm{ep}}$ is calculated as: $\overline{R}_{\mathrm{ep}} := \frac{1}{N_{\mathrm{ep}}} \sum_{k=1}^{N_{\mathrm{ep}}} \sum_{t=1}^{T} \phi(t)$, where N_{ep} is the total number of episodes and T is the total number of time steps.

The final order parameter $\phi(T)$ indicates the level of alignment at the last time step, with the average final order parameter $\overline{\phi}_f$ given by $\overline{\phi}_f := \frac{1}{T_{ep}} \sum_{k=1}^{T} \phi(T)$.

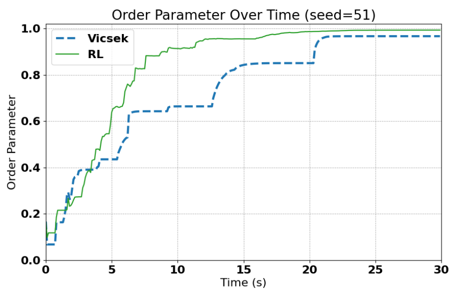
Defined as the proportion of selected neighbors to available neighbors of all agents, the information utilization ratio $\eta_I := \frac{\sum_{i=1}^{N} \sum_j \mathbb{1}(j \in \tilde{N}_i(t) \setminus i)}{\sum_{i=1}^{N} \sum_j \mathbb{1}(j \in N_i(t) \setminus i)}$, where $\tilde{N}_i$ refers to the index set of the selected neighbors and $\mathbb{1}(\cdot)$ denotes an indicator function. The average information utilization ratio: $\overline{\eta}_I := \frac{1}{N_{ep} T} \sum_{k=1}^{N_{ep}} \sum_{t=1}^{T} \eta_I$. Here, $\overline{\eta}_I$ reflects the system's information overhead, where a lower value indicates better performance.

3.2 Simulation Result

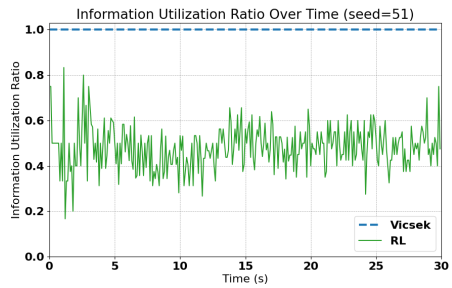
Table 1 compares the average performance of the Vicsek flocking model with the proposed DRL approach (referred to as 'RL') over 200 episodes. The RL approach achieved a slightly higher average episode reward $\overline{R}_{ep}$ and final order parameter $\overline{\phi}_f$ than the Vicsek model, indicating a slight improvement in flocking performance. However, the most notable difference was in the information utilization ratio $\overline{\eta}_I$, where the RL approach used less than half of the available information, reducing computational and communication overhead.

Table 1: Average Performance of Vicsek and RL Approaches (200 Runs)

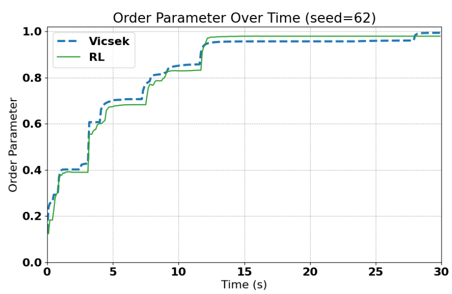
Metric	Vicsek Model	RL Approach
Average Episode Reward ($\overline{R}_{ep}$)	244.4	**252.9**
Final Order Parameter ($\overline{\phi}_f$)	0.9656	**0.9711**
Information Utilization Ratio ($\overline{\eta}_I$)	1.0	**0.4742**



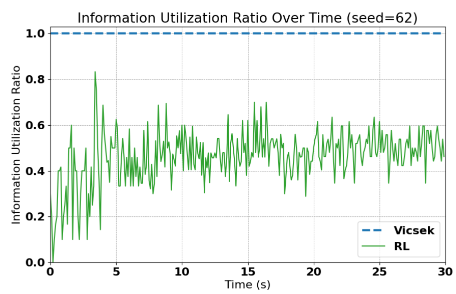
(a) Order parameter



(b) Information utilization ratio



(c) Order parameter



(d) Information utilization ratio

Fig. 1: Order parameter and information utilization ratio

Figure 1 presents the changes in the order parameter and information utilization ratio over time for both approaches in two experiments. In Figs 1b and 1d, the RL method consistently used significantly less information without compromising the flocking behavior as shown in the order parameter levels in Figs 1a and 1c. This indicates that the reduced neighbor information had minimal impact on interaction results, suggesting that the RL method effectively utilized essential information while discarding redundant data.

Despite the stochastic nature of the DRL-based policy, it maintained stable order parameter levels during the episodes, comparable to the deterministic Vicsek model; an agent may occasionally take a large, sudden control input due to the probabilistic sampling of neighbors at each time step. This stability, even in a chaotic multi-agent system, highlights the efficiency of the DRL approach in optimizing flocking behavior with reduced information utilization.

4 Conclusion

This paper proposed an information-efficient flocking method using DRL within the Vicsek model. Our method demonstrated that agents could achieve faster convergence and superior alignment while significantly reducing the amount of neighbor information utilized. This reduction in information overhead highlights the potential for more efficient communication and computation in multi-agent systems. The findings suggest that DRL can autonomously learn effective strategies for neighbor selection, enhancing collective behavior without extensive domain-specific knowledge or assumptions.

Future work could explore applications of this approach in more complex environments, such as those with dynamic obstacles or communication delays as well as more extensive statistical analysis. Additionally, extending the framework to heterogeneous agents with different capabilities and roles could provide further insights into optimizing collective behaviors in diverse scenarios.

References

1. Charbonneau, D., Dornhaus, A.: When doing nothing is something. how task allocation strategies compromise between flexibility, efficiency, and inactive agents. **17** 217–242 Publisher: Springer.
2. Strogatz, S.: Sync: The emerging science of spontaneous order. Publisher: Penguin UK.
3. Centola, D.: The spread of behavior in an online social network experiment. **329**(5996) 1194–1197 Publisher: American Association for the Advancement of Science.
4. Strandburg-Peshkin, A.: Visual sensory networks and effective information transfer in animal groups. **23** 711
5. Ugander, J., Backstrom, L., Marlow, C., Kleinberg, J.: Structural diversity in social contagion. **109**(16) 5962–5966 Publisher: Proceedings of the National Academy of Sciences.
6. Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., Viale, M., Zdravkovic, V.: Interaction ruling animal collective behavior depends on topological rather than metric distance: evidence from a field study. **105**(4) 1232–1237 Publisher: Proceedings of the National Academy of Sciences.
7. Lu, X., Zhang, C., Huang, C., Qin, B.: Research on swarm consistent performance of improved vicsek model with neighbors' degree. **588** 126567
8. Lu, X., Zhang, C., Qin, B.: An improved vicsek model of swarm based on remote neighbors strategy. **587** 126553
9. Martin, S.: Multi-agent flocking under topological interactions. **69** 53–61
10. Zhang, X., Fan, S., Wu, W.: Enhancing synchronization of self-propelled particles via modified rule of fixed number of neighbors. **629** 129203
11. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., Shochet, O.: Novel type of phase transition in a system of self-driven particles. **75**(6) 1226–1229 Publisher: American Physical Society.
12. Bahdanau, D.: Neural machine translation by jointly learning to align and translate
13. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms