The Effect of Network Connectivity on Exploration and Exploitation During Decentralized Collective Learning

Hian Lee Kwa[1,2] 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, roland.bouffanais@uottawa.ca

Introduction

In the field of multi-agent systems (MAS), a well known challenge faced by practitioners is the exploration-exploitation dilemma. This dilemma arises from the fact that the process of gathering information (i.e., exploration) and its usage (i.e., exploitation) tend to be two mutually exclusive activities. In the scenario of collective learning, where a system is tasked with learning the true state of its environment, a strongly exploration-biased system would take excessive amounts of time to come to a final consensus about its environment. Conversely, a strongly exploitation-biased one may wrongly characterize its environment, especially when the agents have to contend with noise (Raoufi et al., 2021). One method to alter the exploration-exploitation balance of a system is through changing an agent's level of connectivity (i.e., changing the number of neighbors with which an agent communicates directly) (Kwa et al., 2022). Indeed, it has been shown in several scenarios that there exists an optimal level of connectivity to maximize a system's performance (Mateo et al., 2019; Kwa et al., 2020, 2021). The work presented here expands on the research previously done by Crosscombe and Lawry (2021) on a decentralized MAS carrying out a collective learning task and further explores the role that agent connectivity plays in regulating a system's transition from exploration to exploitation during such a task.

Method

In Crosscombe and Lawry (2021), the authors describe an environment using a set of n propositional variables $\mathcal{P} = \{p_1, \ldots, p_n\}$, where $p_i \in [0, 1]$, with 0 and 1 representing the false and true states respectively. Each agent assigns a truth value to each of these variables based on its belief, given as $b : \mathcal{P} \to \{0, \frac{1}{2}, 1\}^n$, where $\frac{1}{2}$ signifies that the agent is uncertain of the state of the associated propositional variable. The overall belief of an agent, i, can therefore be represented by the n-tuple, $\langle B_i(p_1), \ldots, B_i(p_n)\rangle$. To facilitate collective learning, the agents are connected to each other using a static topological k-nearest neighbor network. This is essentially a small-world network with the rewiring probability, ρ, set to zero (Watts and Strogatz, 1998). At each time-step, a single random agent pair would be chosen to fuse their beliefs according to the fusion operator first proposed by Crosscombe and Lawry (2017):

$$B_i \odot B_j = \langle B(p_1) \odot B(p_2), \ldots, B(p_n) \odot B(p_n)\rangle, \quad (1)$$

where both agents i and j would adopt the belief $B_i \odot B_j$. At each time-step, all agent are allowed to sample from the environment. To do so, the agent picks a random proposition about which it is uncertain (i.e., such that $B(p_i) = \frac{1}{2}$) to investigate. Upon choosing which proposition to investigate, the environment yields evidence to the agent with a probability r and does not yield any evidence with a probability $1 - r$, where r is known as the evidence ratio. The evidence takes the form of an assertion, $E = \langle \frac{1}{2}, \ldots, S^*(p_i), \ldots, \frac{1}{2}\rangle$, where $S^* : \mathcal{P} \to \{0, 1\}^n$ is the true state of the environment. The agent then updates its belief set using the same operator described in Eq. 1. As the sampling process is affected by noise, the evidence yielded is determined as follows:

$$E(p_i) = \begin{cases} S^*(p_i) : \text{with probability } 1 - \epsilon, \\ 1 - S^*(p_i) : \text{with probability } \epsilon. \end{cases} \quad (2)$$

For further details regarding the information sampling and fusion process, the reader is directed to Crosscombe and Lawry (2017, 2021).

To measure the accuracy of a system, the average error is calculated at the end of each run by finding the mean Hamming distance between the agents' beliefs and the ground truth. In addition, to measure the level of exploration of the MAS, the number of unique beliefs sets maintained by the system are tracked across the duration of the simulation, with higher number of belief sets signifying a greater amount of exploratory activity.

In the simulations performed, 100 agents are used to characterize an environment consisting of $n = 1,000$ propositions. For each set of parameters, 50 trials were carried out to obtain the average result. These trials lasted for a maximum of $10,000$ iterations or until the system had converged on a consensus belief.

Results

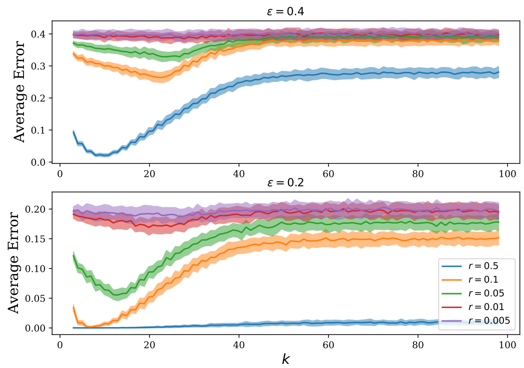

Figure 1: Average error of a 100 agent system characterizing a $n = 1,000$ proposition environment with noise levels of $\epsilon = 0.4$ (top) and $\epsilon = 0.2$ (bottom). This was measured at the end of each simulation run and averaged over 50 runs.

Fig. 1 illustrates an optimum level of connectivity, k, that minimizes the average error of the agents' beliefs and is present at different noise levels. Similar to the findings of Crosscombe and Lawry (2021), this only occurs when the evidence rate, r, is sufficiently high (i.e., for $r > 0.01$). The figure also shows that this optimum decreases with increasing evidence rates as well as with decreasing noise levels.

These findings can be explained through the analysis of Fig. 2; when both r and k are high, after the system's initial exploration, it converges very quickly to a consensus. This is characterized by a high initial number of belief sets maintained by the system followed by a sharp decrease that signifies a rapid switch from exploration to exploitation. However, in the presence of environmental noise, the system ultimately converges on a single belief set with a high level of error when compared to the ground truth (i.e., the system has exploited a poor source of information). High average errors are also observed in low k systems. Unlike the high k cases, these errors stem from the system maintaining elevated levels of exploration throughout the simulation. As such, the exploration-exploitation transition does not occur and the system is unable to converge on a consensus within the time limit. This high number of belief sets maintained by the system at the end of each simulation run is what leads to a high level of average error. In contrast, when the system operates at the optimal level of connectivity, it is able to maintain a high number of belief sets, and thereby exploration, for a sufficient period of time and converges on a consensus just before the maximum allowable time elapses.

At low evidence rates, the system gathers less information from the environment. This reduces the effectiveness of the exploration process, but also allows the agents to keep their individual belief sets constant for longer periods of time, thus facilitating the consolidation of these sets. This reduction in the number of belief sets present in the system suggests that a small amount of exploitation is being performed by the MAS. However, due to the scarcity of information and the presence of noise, systems using low levels of connectivity are unable to fully transition from exploration to exploitation within the allocated time. This is evidenced by Fig. 2 that shows systems using $k = 10$ and $k = 20$ networks are unable to converge on a consensus, thereby leading to high levels of error. The figure also shows that while systems using high values of k are ultimately able to settle on a consensus, they are still unable to gather enough accurate information and therefore converge on a single shared belief that does not accurately reflect the ground truth.

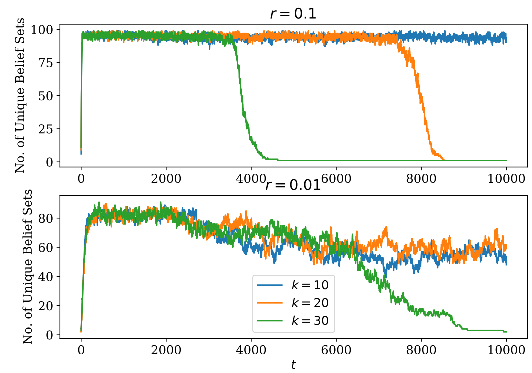

Figure 2: Number of belief sets maintained at time t for various levels of connectivity, k, during a single simulation run. Simulations were run with a noise level of $\epsilon = 0.4$ and evidence rates of $r = 0.1$ (top) and $r = 0.01$ (bottom).

Discussion

In this work, we have shown how an MAS network's connectivity degree can be used to regulate the exploration-exploitation transition in a decentralized learning task. The speed of this transition impacts the overall performance of a swarm; while using a high level of connectivity quickens this transition and reduces convergence time at the cost of reduced system accuracy, low levels of connectivity prolongs the convergence time but results in a more accurate characterization of the environment. The trade-off between convergence speed and accurate environment characterization leads to the presence of an optimum level of connectivity that minimizes the consensus error. Such optima are also found in other systems such as target tracking and leader-follower MAS. A limitation of this work is that agent movement and imperfect communications are not considered; agents 'teleport' around the environment to sample different propositions while the communications network remained constant. In addition, only one agent pair is selected for belief fusion at each time-step; the system still retains some centralization. Future work should include the modelling of agent movement and the study of environment sampling rates to prevent the spatial correlation of beliefs. Methods using dynamic communications networks that permit individual agents to decide when fuse beliefs should also be developed. These studies would allow the findings to be better implemented in truly decentralized robotic systems.

Acknowledgements

We would like to thank Dr. Michael Crosscombe for the fruitful discussions and his comments on our manuscript. This research is funded by Thales Solutions Asia Pte Ltd, under the Singapore Economic Development Board IPP.

References

Crosscombe, M. and Lawry, J. (2017). Exploiting vagueness for multi-agent consensus. *Studies in Computational Intelligence*, 670:67–78.

Crosscombe, M. and Lawry, J. (2021). The impact of network connectivity on collective learning. In *Proceedings of the 15th International Symposium on Distributed Autonomous Robotics Systems (DARS21)*. Springer Cham.

Kwa, H. L., Kit, J. L., and Bouffanais, R. (2020). Optimal swarm strategy for dynamic target search and tracking. In *Proc. of the 19th International Conference 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.

Kwa, H. L., Kit, J. L., and Bouffanais, R. (2022). Balancing collective exploration and exploitation in multi-agent and multi-robot systems: A Review. *Frontiers in Robotics and AI*, 8(771520).

Mateo, D., Horsevad, N., Hassani, V., Chamanbaz, M., and Bouffanais, R. (2019). Optimal network topology for responsive collective behavior. *Science Advances*, 5(4):eaau0999.

Raoufi, M., Hamann, H., and Romanczuk, P. (2021). Speed-vs-accuracy tradeoff in collective estimation: An adaptive exploration-exploitation case. In *2021 International Symposium on Multi-Robot and Multi-Agent Systems (MRS)*, pages 47–55, Cambridge, United Kingdom. IEEE.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature*, 393(6684):440–442.