Emergent Specialization in Swarm Systems

Ling Li, Alcherio Martinoli, and Yaser S. Abu-Mostafa

California Institute of Technology, Pasadena, CA 91125, USA

Abstract. Distributed learning is the learning process of multiple autonomous agents in a varying environment, where each agent has only partial information about the global task. In this paper, we investigate the influence of different reinforcement signals (local and global) and team diversity (homogeneous and heterogeneous agents) on the learned solutions. We compare the learned solutions with those obtained by systematic search in a simple case study in which pairs of agents have to collaborate in order to solve the task without any explicit communication. The results show that policies which allow teammates to specialize find an adequate diversity of the team and, in general, achieve similar or better performances than policies which force homogeneity. However, in this specific case study, the achieved team performances appear to be independent of the locality or globality of the reinforcement signal.

1 Introduction

Swarms of relatively simple autonomous agents can exhibit complex behavior which appears to transcend the individual ability of the agents. Perhaps the most striking examples are from nature: social insect colonies are able to build sophisticated structures and regulate the activities of millions of individuals by endowing each individual with simple rules based on local perception. Swarm intelligence is an innovative computational and behavioral metaphor for solving distributed problems that takes its inspiration from the behavior of social insects [1] and swarming, flocking, herding, and shoaling phenomena in vertebrates [2].

Artificial swarm systems based on swarm intelligence are truly distributed, self-organized, and inherently scalable since there is no global control or communication. The agents are designed to be simple and interchangeable, and may be dynamically added or removed without explicit reorganization, making the collective system highly flexible and fault tolerant. One of the domains in which the swarm intelligence approach has been successfully applied is collective mobile robotics. In this paper, we will present a specific case study in this domain previously investigated using hand-coded solutions only. We will show that distributed learning represents an effective method for automatically selecting individual control parameters, which in turn influence the team performance.

Martinoli and Mondada [3] and successively Ijspeert et al. [4] investigated collaboration in groups of reactive, non-communicating robots engaged in a stick pulling experiment (see Fig. 1). In this experiment, a team of robots search a circular arena and pull sticks out of holes in the ground. The length of a stick

H. Yin et al. (Eds.): IDEAL 2002, LNCS 2412, pp. 261–266, 2002.
© Springer-Verlag Berlin Heidelberg 2002

Fig. 1. Physical set-up for the stick pulling experiment.

has been chosen so that a single robot is not capable of pulling a stick out on its own. Collaboration between two robots is necessary for solving this task. Each robot is characterized by a gripping time parameter (GTP), which is the length of time that a robot waits for the help of another robot while holding a stick. Two cases can occur: either a second robot helps the first one (we define this as a successful collaboration) or the GTP expires before any other robot can help and the first robot resumes the search for sticks in the arena. The specific values of GTPs play a crucial role in the overall collaboration rate (defined as the number of successful collaborations per unit time), which is the metric adopted in both previous papers as well as in this one for measuring the team performance.

In addition to experiments performed using real robots and sensor-based, embodied simulations, Ijspeert et al. proposed a microscopic model which delivered both qualitatively and quantitatively accurate predictions. In this paper, we exploit this result by integrating our distributed learning algorithms into their microscopic model. As a consequence, although we have not tested our learning algorithms using real robots or realistic simulations, we believe that their validity is not limited to abstract agents.

Alternatively, optimal control parameters of a swarm system can be computed with the help of macroscopic models [5]. However, macroscopic models have two major drawbacks in comparison with a machine learning method. First, quantitatively correct macroscopic models based uniquely on features of the individual agents are not always trivial to devise, in particular when agent-to-agent and agent-to-environment interactions are more complicated than simple elastic bounces. Second, these models intrinsically assume that a certain number of agents can be clustered in a caste which in turn is represented by a set of differential or difference equations. In our machine learning approach, we do not assume the existence of any caste (each agent in principle can be different from any other agent) and, if clusters of specialists arise, these are due to the learning process rather than to a priori established categories.

Finally, depending on the size of the search space and the number of agents, a systematic investigation of the optimal individual control parameters (such as that conducted in [4]) could be prohibitively time consuming even allowing for the fact that in this particular case study, microscopic models have been proven to be four or five orders of magnitude faster than real robot experiments.

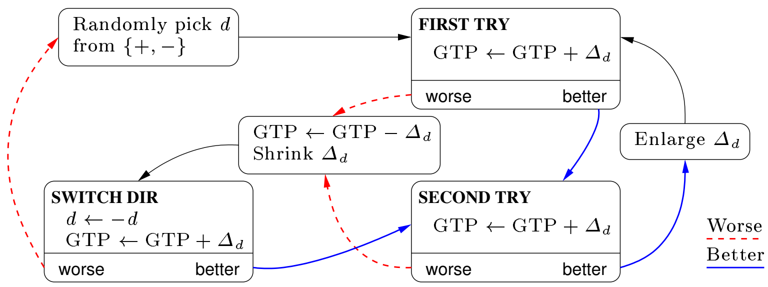


Fig. 2. State graph of the learning algorithm. GTP is adjusted according to the current state and performance change (better or worse). $\varDelta_d$ is the search step in direction d.

2 Learning Algorithm

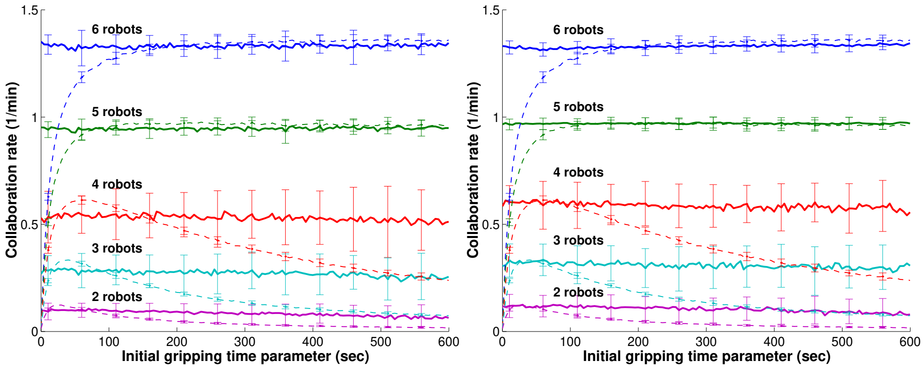
The major challenge in designing a distributed learning algorithm able to tune individual parameters so that the team performance is optimized is to solve the credit assignment problem [6] in a system where the information available locally is incomplete and noisy. In our specific case study, the optimal GTPs depend on the number of other robots working together, the environmental characteristics (e.g., number of sticks), and other robots' GTPs. Without global communication and perception, all these conditions can only be estimated by an individual agent.

The learning principle we propose in this paper is intuitively similar to what a human being would do in a partially (locally) known environment (see Fig. 2). The agent first tries to change its GTP in a randomly chosen direction. After the change, the agent maintains that GTP for a small period of time and monitors the performance improvement. If the improvement is positive, the agent will continue in that direction; if negative, the agent will undo the last change and try the other direction. The search step $\varDelta_d$ can vary between $[\varDelta_{\min}, \varDelta_{\max}]$ and is adjusted according to the success of the previous try.

Note that in this diagram the change of GTP is linear, meaning that in k periods of time, the maximal change is $k\varDelta_{\max}$. We call this linear way of adjusting the GTP the "$\varDelta$-method." However, since the collaboration rate is much less sensitive to changes in the GTP when the GTP is large, a more effective search method for large GTPs is to use a search step proportional to the absolute value of the GTP. We call this alternative way of tuning the GTP the "%-method." Since a specific method is more effective than the other in a given part of the search space, in the following we use a hybrid learning algorithm which alternates sequentially both methods.

3 Experimental Results

We ran stick pulling experiments using two to six robots and four sticks. Each run, characterized by different initial GTPs, was repeated 100 times. For each experiment, we first conducted a 1600 min learning phase, during which robots



(a) homogeneous (b) heterogeneous

Fig. 3. The collaboration rates with and without learning for varying number of robots. Robots were initialized with the same GTP. The dashed curves in both plots represent the collaboration rates obtained by systematic search and homogeneous teams while the solid lines show those after learning with (a) global reinforcement and homogeneous teams; or (b) local reinforcement and heterogeneous teams.

could iteratively adapt their GTPs. Then a 1600 min test phase was conducted to measure the performance of the learned solutions. In all of the following plots, the error bars represent the standard deviation over different repetitions.

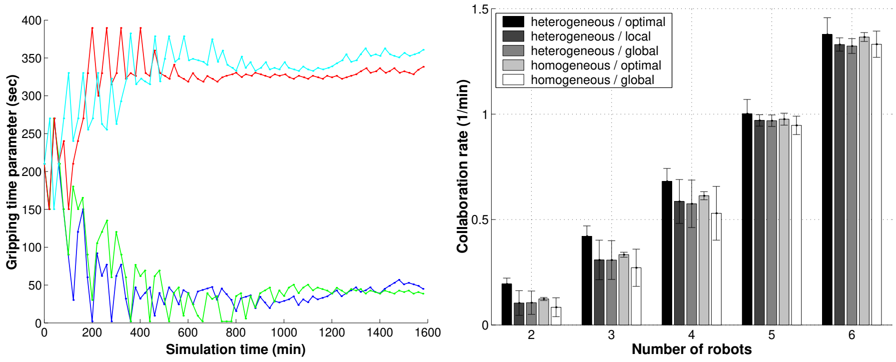
3.1 Homogeneous Teams

We first investigated homogeneous teams, meaning that all the robots were characterized by the same GTP. We compared systematic hand-coded results with learned solutions using global reinforcement[1].

Figure 3(a) and Fig. 4(b) compare the collaboration rates obtained with and without learning. The collaboration rate consistently achieves the same level, independent of the initial GTP and is above the average collaboration rate obtained with fixed GTPs. With teams of five or six robots, the learning algorithm achieves the optimal average collaboration rate (as obtained in the systematic search) for each initial GTP. This is not the case for teams of two to four robots, although the maximal collaboration rate obtained in the systematic study is within one standard deviation of that reached with learning.

Further studies are needed to understand the limitations of the proposed learning algorithm. In particular we will closely investigate the role of noise affecting the reinforcement on the convergence towards a global optimum.

[1] While local reinforcement is more realistic for a swarm system, global reinforcement, which usually implies physically a supervisor that measures and broadcasts the team performance to the individual agents, provides an interesting term of comparison.



(a) 4 robots formed 2 clusters

(b) performance

Fig. 4. Specialization and performance. (a) 4 robots using the Δ-method had 210 sec as the initial GTP. At the end of the simulation, they formed two clusters, one characterized by a large GTP and the other by a small GTP. (b) Performances obtained with a systematic search and after learning under different reinforcement and team diversity.

3.2 Heterogeneous Teams

For each robot, the local reinforcement is the rate of successful collaborations (regardless of whether a robot was the first or the second in gripping the stick). Figure 3(b) shows the performance after learning outperforms that in the homogeneous teams. In most cases the robots became specialized after learning. Figure 4(a) shows the GTP curves of four robots in a single run. All robots started with the same initial GTP. As the experiment progressed, their GTPs diversified and formed two clusters. Our results are consistent with the results of systematic experiments reported in [4] which showed that, when the number of robots is less than or equal to the number of sticks, specialization is helpful.

The increase in the performance lies essentially in the heterogeneity of the solution, rather than in the learned values themselves. Indeed, a systematic study of all possible GTP combinations (see Fig. 4(b)) has shown that similar problems of noise and convergence found in homogeneous team also exist in this case—learned solutions perform slightly worse than those found by systematic search[2].

For the sake of completeness, following [7] we investigated heterogeneous teams combined with global reinforcement. Once again, probably due to the high alignment of individual and team performances, we did not obtain any significant difference in the quality of the learned solution as a function of the reinforcement type (see Fig. 4(b)).

[2] For heterogeneous teams, we tested all GTP combinations where GTP (in seconds) is from the set $\{5k\}_{k=1}^{20} \cup \{10k\}_{k=11}^{15} \cup \{175, 200, 250, 300, 400, 500, 750, 1000\}$.

4 Conclusions

In this paper, we investigated several distributed learning policies based on different reinforcement signals (global and local) and different team diversity (homogeneous and heterogeneous). The discussion was based on a faithful microscopic probabilistic model previously developed for investigating the team performance in a concrete case study in collective robotics (the stick pulling experiment). The results show that policies which allow teammates to specialize find an adequate diversity of the team as a function of the task constraints and, in general, achieve similar or better performances than policies which force homogeneity.

The obtained specialization is an interesting result, since we never explicitly rewarded diversity nor there was any explicit communication between agents. In this specific case study, the achieved team performances appear to be independent of the locality or globality of the reinforcement signal, probably due to the high alignment between both forms of reinforcement.

Acknowledgements. We would like to acknowledge Lavanya Reddy and Eric Tuttle for having implemented a first version of the Δ-method we are using in this paper.

This work was supported by the Caltech Center for Neuromorphic Systems Engineering under NSF Cooperative Agreement EEC-9402726.

References

1. Bonabeau, E., Dorigo, M., Théraulaz, G.: Swarm Intelligence: From Natural to Artificial Systems. Oxford University Press, New York (1999)
2. Parrish, J.K., Hamner, W.M., eds.: Animal Groups in Three Dimensions. Cambridge University Press, New York (1997)
3. Martinoli, A., Mondada, F.: Collective and cooperative group behaviours: Biologically inspired experiments in robotics. In Khatib, O., Salisbury, J.K., eds.: Proceedings of the Fourth International Symposium on Experimental Robotics (1995). Lecture Notes in Control and Information Sciences, Vol. 223. Springer-Verlag, Berlin (1997) 3–10
4. Ijspeert, A.J., Martinoli, A., Billard, A., Gambardella, L.M.: Collaboration through the exploitation of local interactions in autonomous collective robotics: The stick pulling experiment. Autonomous Robots **11** (2001) 149–171
5. Lerman, K., Galstyan, A., Martinoli, A., Ijspeert, A.J.: A macroscopic analytical model of collaboration in distributed robotic systems. Artificial Life **7** (2001) 375–393
6. Versino, C., Gambardella, L.M.: Learning real team solutions. In Weiß, G., ed.: Distributed Artificial Intelligence Meets Machine Learning: Learning in Multi-Agent Environments. Lecture Notes in Artificial Intelligence, Vol. 1221. Springer-Verlag, Berlin (1997) 40–61
7. Murciano, A., Millán, J. del R., Zamora, J.: Specialization in multi-agent systems through learning. Biological Cybernetics **76** (1997) 375–382