Distributed Learning in Swarm Systems

Ling Li, Alcherio Martinoli, Yaser Abu-Mostafa
Learning Systems Group, CNSE, Caltech
http://www.work.caltech.edu

Motivation and Aims

Natural systems consisting of many agents, such as ants, wasps, and termites, appear to have the ability to transcend the constituent individual agents. Scalability, flexibility and robustness are three main advantages for such swarm intelligence systems.

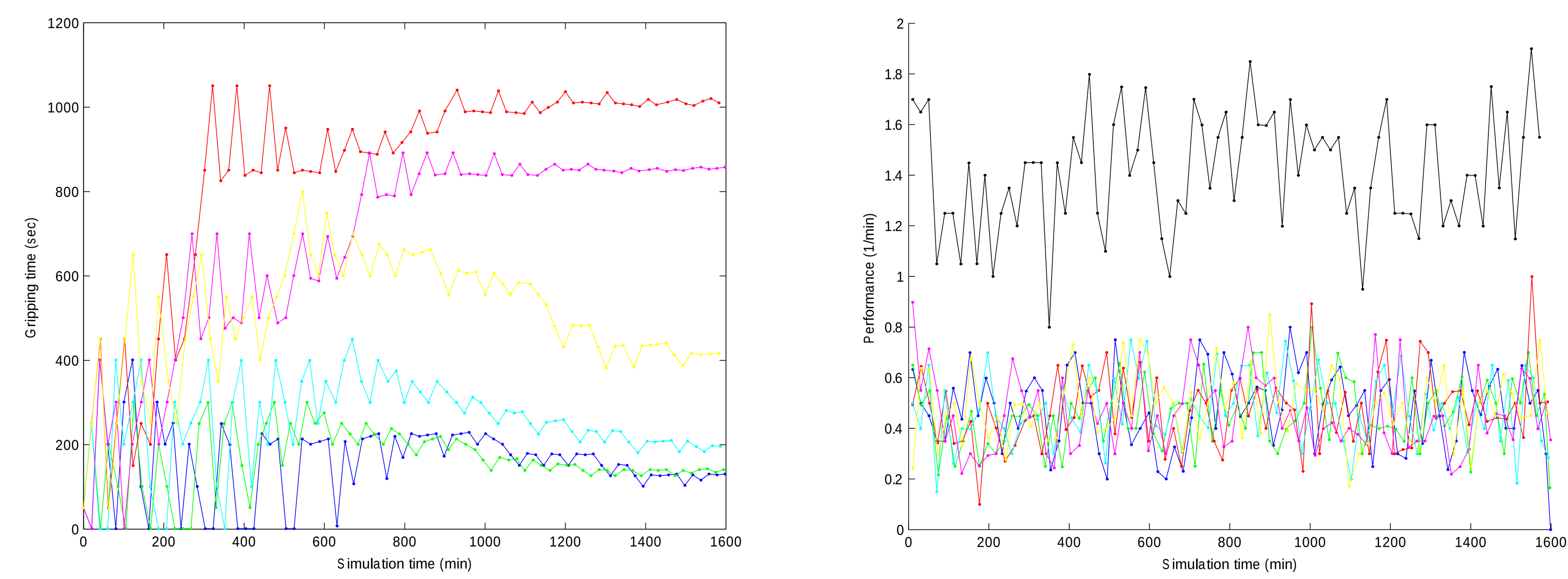
Our aim is to apply the principles inspired from these natural systems to distributed problems, such as the control of a swarm of robots. We would also like to investigate the role of individual learning capabilities on the emerging collective behavior.

Stick-pulling Problem

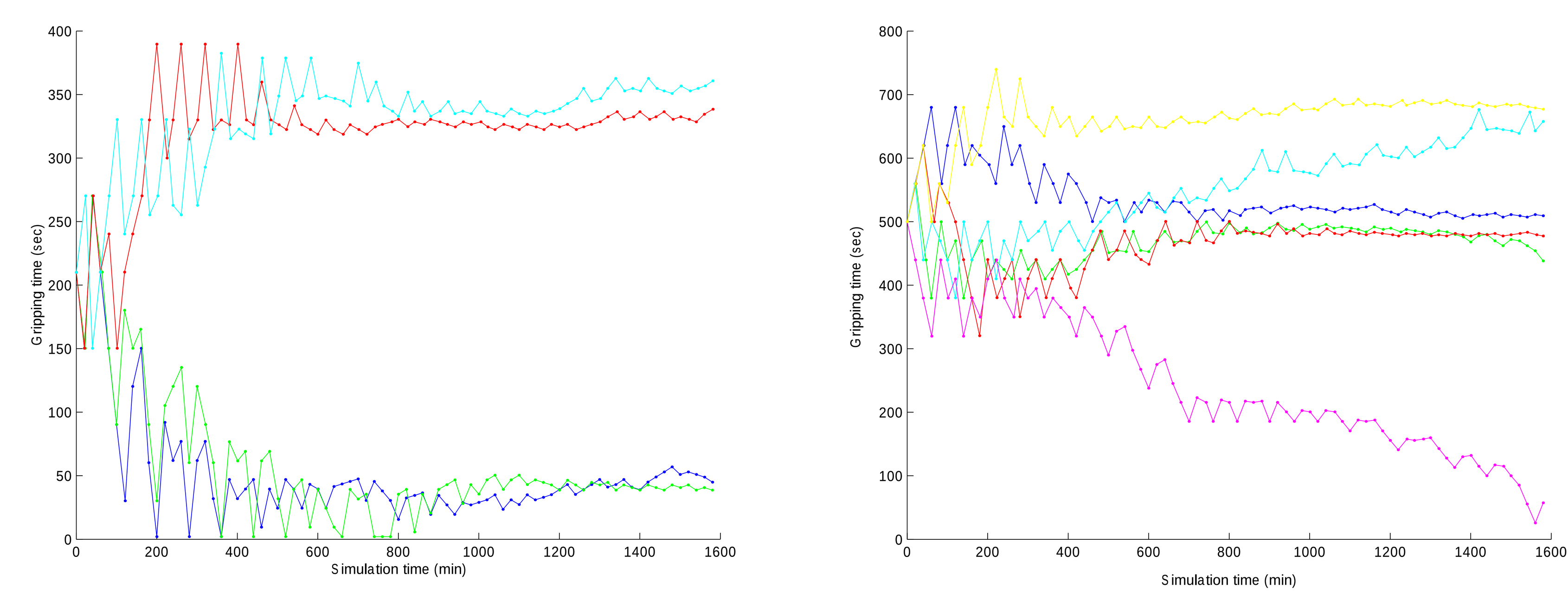
We looked at the stick-pulling problem where multiple robots in an arena worked on a task (pulling sticks out of the ground) that cannot be done without collaboration. The learning task here is to find the optimal gripping time parameter (GTP) for each robot, in order to maximize the group performance.

By different means of communication, robots can use public or private knowledge; by different feedback from the environment, the learning can be conducted with group reinforcement or individual reinforcement; by forcing all the robots to be the same or not, the parameters setting could be homogeneous or heterogeneous. We try to analyze and design different learning algorithms for different combinations.

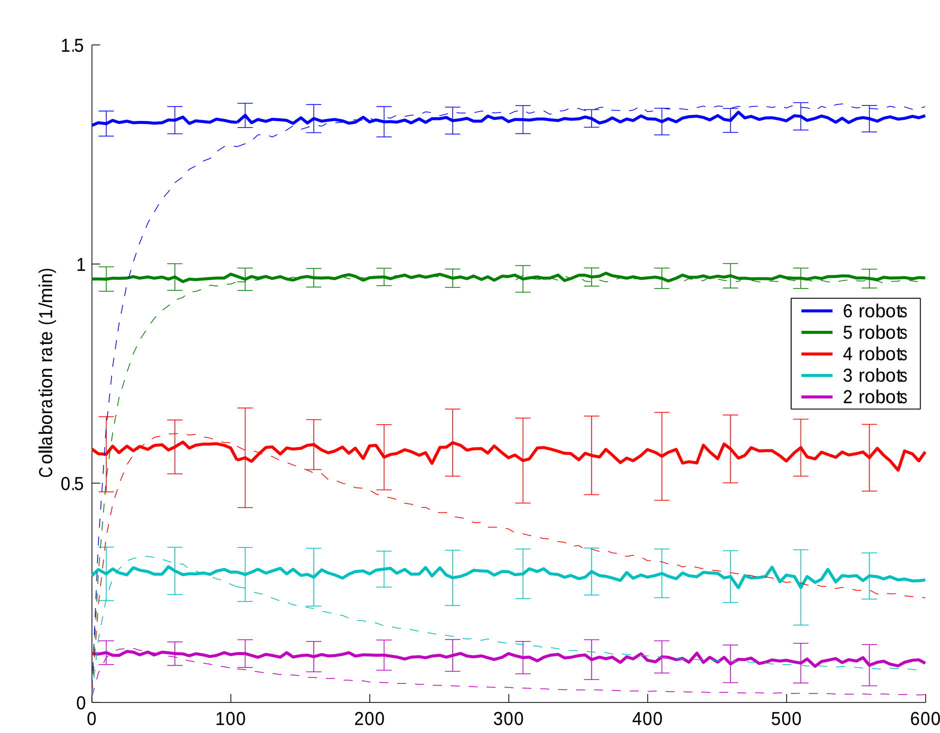
The stick-pulling problem. In the figure, 6 robots with gripper turrets and proximity sensors are trying to pull sticks out of the ground. The stick is long so that one robot can not pull it out by itself. They need collaboration. The gripping time is the time a robot will wait while holding a stick.



The noisy environment. The initial GTPs were set to 50s. (Left) GTP curves of individual robots. (Right) The overall and individual performance (collaboration rate) curves.



Examples of specialization. (Left) 4 robots had 210s as the initial gripping time. At the end of the simulation, they formed two groups, one with large GTP and the other with small GTP. (Right) 6 robots had 500s as the initial GTP. Three groups were formed at the end of the simulation.



The performance (collaboration rate) with learning. Different colors represent experiments with different number of robots. Robots were initially given a gripping time. With learning, they adjusted their gripping time and achieved a higher performance. Error bars are standard deviations of performance over 50 runs. Dashed curves are performance without learning.

Approach

Instead of experimenting with real robots, we used a probabilistic model in simulation. The probabilistic model describes the experiment as a series of stochastic events with probabilities based on simple geometrical considerations and systematic experiments with one or two real robots, and is quite fast in simulation.

The learning was conducted under individual reinforcement. Several methods, such as adaptive line search and Q-learning, were used to find the optimal gripping time.

Difficulties

- Each agent can only sense a small part of the overall system, and the price for full communication is very high. Thus, the learning of each agent is usually conducted with only partial information.
- The environment is noisy. Robots move around randomly and their GTPs change separately. Thus the individual performance changes a lot over time.

Results with No Communication

Below in the middle is a plot of performance v.s. initial gripping time. The augmented solid curves show that the performance does increase with learning.

The results also showed that after learning the robots usually became specialized. This is quite interesting since we never incorporated preference for specialization in the learning algorithm, and there was no communication among the robots. Some previous research showed with a systematic study that under certain constraints there was an advantage in being specialized.

Future Work

Currently we are studying the case with local communication, which intuitively would help robots have more precise and thorough knowledge about the environment. This may lead to a faster and more stable learning.