Group Behavior and Group Learning

Maja J Mataric

MIT AI Lab
545 Technology Square #721
Cambridge, MA 02139
E-mail: maja@ai.mit.edu

1 Synthesis and Analysis of Group Behavior

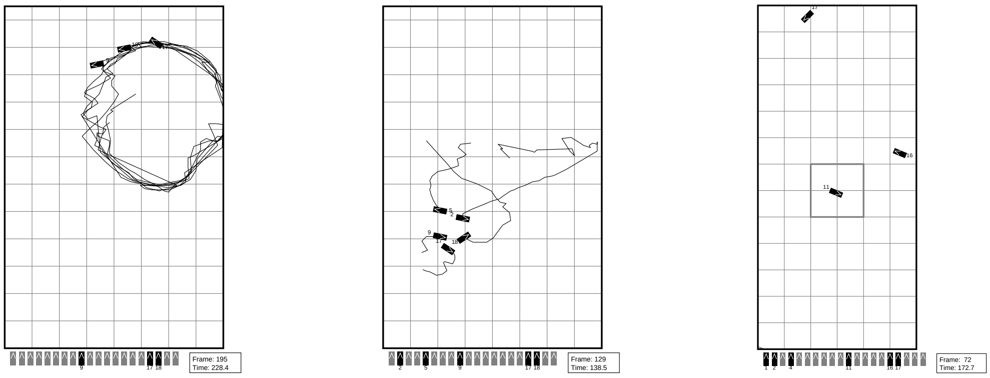
Figure 1. The mobile robots used to demonstrate and verify our group behavior and learning work. These robots demonstrated group avoidance, following, aggregation, dispersion, flocking, wandering, foraging, docking, and learning to forage.

Our work is based on the belief that intelligent collective behavior in a decentralized system results from *local interactions* based on simple rules. **Basic behaviors** are proposed as a methodology for structuring those rules through a principled process of synthesis and evaluation. A behavior is a control law that achieves and maintains some goal.

We postulate that, for each domain, a set of behaviors can be found that are **basic** in that they are required for generating other behaviors, as well as being a minimal set the agent needs to reach its goal repertoire. The process of choosing the set of basic behaviors for a domain is dually constrained. From the bottom up, the process is constrained by the agent and environment dynamics. From the top down, the process is constrained by the repertoire of the agent's goals.

An effective set of basic behaviors in the spatial domain should enable the agents to employ a variety of flexible strategies for puck manipulation, collection, and distribution. The effectiveness of such strategies depends on maximizing synergy between agents: achieving the necessary goals while minimizing inter–agent interference.

The following set of basic behaviors is proposed: *avoidance*, *following*, *aggregation*, *dispersion*, *homing*, and *wandering*. According to our definition, the above behavior set is *minimal*



Following Homing Dispersing

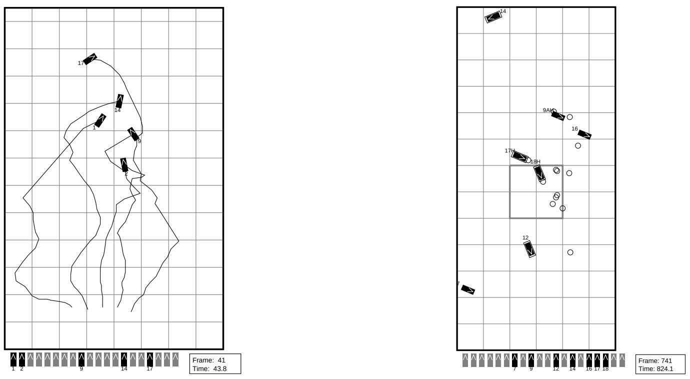
Figure 2. Examples of real robot data for three basic behaviors: following, homing, and dispersing. The robots, physically 12 inches long, are scaled down and plotted as black rectangles, with white arrows indicating their heading. The the row of rectangles at the bottom shows the robots that were used in the experiment. Boxes on the lower right indicate frame numbers and the elapsed time in seconds for each of the runs.

or *basic* in that its members are not further reducible to each other, and they are sufficient for achieving our set of pre–specified goals. A number of other utility behaviors can be a part of an agent's repertoire, such as *grasping* and *dropping*, the only other behaviors used in this work.

The basic behavior set is evaluated by giving a formal specification of each of the behaviors, and comparing the collection of those specifications to a formal specification of the set of global tasks required for the group. We have provided detailed specifications and algorithms for each of the basic behaviors, implemented and evaluated them based on the following criteria: *repeatability*, *stability*, *robustness*, and *scalability*.

The above criteria were applied to the data obtained by running at least 50 trials of each basic behavior on a collection of over 20 physical mobile robots equipped with local sensors and local control (figure 1). Each of the robots is equipped with a suite of infra–red sensors for collision avoidance and puck detection, micro switches and bump sensors for contact detection, and radios and sonars for triangulating their position relative to two stationary beacons, and broadcasting word–sized messages within a limited radius. The basic behaviors, each consisting of one rule or a small set of simple rules, generated robust group behaviors that met the prespecified evaluation criteria. A small subset of the data is shown in figure 2.

Basic behaviors are intended as building blocks for achieving higher–level goals and can be embedded in an architecture that allows two types of combination: direct (by summation) and temporal (by switching). A simple and robust *flocking* behavior was generated by summing the



Flocking   Foraging

Figure 3. Examples of robot data for two different composite group behaviors: flocking and foraging. The foraging behavior of 7 robots is shown after 13.7 minutes of running; collected pucks are in the box.

outputs of *avoidance*, *aggregation*, *wandering*, and *homing*. A more complex *foraging* behavior was implemented by switching between *avoidance*, *dispersion*, *following*, *homing*, *wandering* in search for pucks, and puck manipulation. Figure 3 demonstrates some robot data from the two composite behaviors.

In addition to empirical testing of the behaviors and their combinations, we compared our methodology to a centralized, "total knowledge" approach. The experimental results showed that the simple, fully distributed strategies, applied to dispersion and aggregation tasks, converged only a constant factor slower than the centralized approach (Matarić 1994*a*).

2 Learning in Complex Group Environments

We developed a methodology for automatically generating higher–level behaviors by having the agents learn through their interactions with the world and other agents by unsupervised **reinforcement learning** (RL). RL has been successfully applied to a variety of domains where the agent–environment interaction can be described as a Markov Decision Process (MDP). That assumption does not directly apply to the stochastic, noisy, and uncertain multi–agent environments addressed in this work. The traditional formulation of RL problems as states, actions, and reinforcement required a reformulation in order to be applied to our domain. Instead of using actions, our system learns at the level of *basic behaviors* that hide low–level control details, and are more general and robust. With actions replaced with behaviors, states can be replaced with *conditions*, the necessary and sufficient subsets of state required for triggering the behavior set. Conditions are many fewer than states, so their use diminishes the agent's learning space and speeds up any RL algorithm.

In addition to the use of behaviors and conditions, we used two ways of shaping the reinforcement function to aid the learner in the nondeterministic, noisy, and dynamic environment. We introduced *heterogeneous reward functions*, which partition the task into subgoals, thus providing more immediate reinforcement. Within a single behavior (i.e., a single goal), *progress estimators* are introduced as functions associated with particular conditions that provided some metric of the learner's performance. Progress estimators decrease the learner's sensitivity to noise, minimize thrashing, and minimize fortuitous rewards by correlating some domain knowledge about progress with appropriate behaviors the agent has taken in the past.

The proposed RL formulation was validated on the task of learning to forage. The behavior space included the foraging subset of basic behaviors, augmented with *grasping*, *dropping*, and *resting*. The condition set was reduced to the power set of: *have-puck?*, *at-home?*, *night-time?*, and *near-intruder?*.

Three learning algorithms were tested; the first was standard RL Q-learning, while the other two simply summed the reinforcement received over time. Q-learning received a reward whenever a robot dropped a puck at home. The second algorithm was based on the reinforcement received from heterogeneous reward functions. The third algorithm used reinforcement both from the heterogeneous reward functions and from two progress estimators (intruder-avoidance and homing) found to be sufficient for making the given learning task possible and for consistent and complete learning performance. The absence of either one disabled the robots from learning the complete policy (Matarić 1994*b*). The performance of each of the three algorithms was averaged over 20 trials. The analysis of the learning performance showed that the parts that were not learned by the first two algorithms relied on the progress estimators and were successfully learned in by third.

3 Summary

This work is intended as a foundation in a continuing effort toward studying increasingly more complex behavior, and through it, more complex intelligence. The work on basic behaviors distills a general approach to control, planning, and learning. The work also empirically demonstrates some challenging problems and offers some effective solutions to group behavior and learning.

References

Matarić, M. J. (1994*a*), Interaction and Intelligent Behavior, PhD thesis, MIT.

Matarić, M. J. (1994*b*), Reward Functions for Accelerated Learning, *in* 'Proceedings, Eleventh International Conference on Machine Learning (ML-94)'.