Machine Learning Methods for Multi Robot Navigation (Doctoral Consortium)

Julio Godoy
Department of Computer Science and Engineering
University of Minnesota
godoy@cs.umn.edu

ABSTRACT

In multi robot navigation, robots need to move towards their goal positions while adapting their paths to account for potential collisions with other robots and static obstacles. Existing methods compute motions that are optimal locally but do not account for the motions of the other robots, producing inefficient global motions when many robots move in a crowded space. In my research approach, each robot uses online machine learning techniques to adapt dynamically its behavior to the local conditions. The approach is highly scalable because each robot makes its own decisions on how to move. Experimental results obtained in simulation under different conditions show that the robots reach their destinations faster using motions that are more energy efficient.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents*

Keywords

multi agent navigation; online learning

1. INTRODUCTION

Real-time goal-directed navigation of multiple robots in crowded environments, where each one of the robots can only observe its nearby robots, has important applications in many domains such as swarm robotics, planning for evacuation, and traffic engineering. This navigation problem is challenging because robots have conflicting constraints. On one hand, they need to reach their goals as soon as possible while avoiding collisions with each other and the static obstacles present in the environment. On the other hand, due to the presence of many robots and the real-time constraints, robots need to compute their motions independently of each other and in a decentralized manner instead of planning in a joint configuration space.

A recently introduced decentralized technique for real-time multi robot navigation, ORCA [6] guarantees collision-free motion for the robots. Although ORCA generates locally efficient motion for each robot, the overall behavior of the robots can be far from efficient; actions that are locally optimal for one robot are not necessarily optimal for the entire group of robots. Consider, for example, the two groups of robots in Fig. 1 that try to pass to the other side of a narrow doorway. Here, congestion arises around the doorway that causes long delays and an inefficient crowd motion. If, instead, robots coming from the back were able to stop or temporarily move away from the doorway, the congestion would resolve and the global motion would be much more efficient.

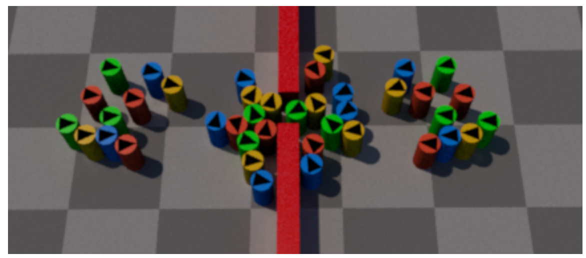

Figure 1: Congestion when two groups of robots pass through a narrow doorway in opposite directions.

My thesis research focuses on applying *online* methods for planning and learning that can be completely distributed and require no communication among the robots. My work aims at enabling robots to make intelligent motion decisions that will take them to their goals faster, for example, to efficiently perform critical tasks such as search and rescue operations or evacuating a building. Online approaches are suitable for dynamic environments, as robots must be able to quickly adapt their behaviors to changes in their surroundings. In contrast, offline approaches are not suitable when quick response time and robustness to dynamic environments are desired, as they require a long training phase and the prediction of all potential scenarios the robot might encounter. In addition, the computational complexity of centralized offline learning methods becomes prohibitively high as the number of robots increases. Because of this, current literature on multi robot learning focuses on scenarios with only a few robots and sparse interactions.

2. CONTRIBUTIONS

I have proposed two methods for improving the global motion of the robots: a learning-based and a planning-based method. In both methods, I extended the range of motions of the robots to increase the diversity of behaviors they can

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

exhibit, beyond the myopic goal-oriented motion commonly considered in the literature [6].

As a first step in developing online learning approaches, I proposed an action selection technique called ϵ-UCB [3], inspired by the principles of well-known action selection techniques, ϵ-greedy and Upper Confidence Bounds (UCB). ϵ-greedy selects either the best or a random action while UCB samples them proportionally to the upper bound of the estimated value of their rewards. In ϵ-UCB, I formulated the problem of selecting the best motion at each timestep as an action selection problem in a multi armed bandit setting. In this formulation, the challenge is to carefully balance action exploration and exploitation. ϵ-UCB exploits the best action in a greedy fashion and performs biased exploration using a version of UCB more suited to non-stationary domains (by using a moving window of the history of the rewards). Combined with a reward function that considers both goal-oriented motion and robot-neighborhood interactions, ϵ-UCB allows robots to dynamically adapt their motion to their local conditions (i.e., move back or sideways when goal-oriented motion is constrained). This indirectly improves the global efficiency of the motions of all robots, allowing them to reach their destination faster.

The realization that robots need to adapt the amount of exploration to their local conditions drove me to further improve this approach. I proposed a novel and general framework for incorporating learning methods in multi robot navigation, the ALAN framework (Adaptive Learning for Agent Navigation) and a new context-aware action selection technique. This context-aware approach improves over ϵ-UCB by introducing game-theoretic elements, considering the local context of the robot to strategically adapt the amount of exploration performed. Robots using the ALAN learning framework and the context-aware technique take advantage of pure goal-oriented motion when they are able to, and perform biased exploration when this motion is constrained. This enables the entire set of robots to reach their destinations faster, scaling to different environments and number of agents, and outperforming ϵ-UCB and other existing action selection techniques [4].

On the planning side, I proposed an anytime local approach to plan the motions of the robots in a decentralized manner, by adapting the Hindsight optimization technique [1] in a progressive manner. I called this method Progressive Hindsight Optimization (PHOP) [2]. PHOP reduces the uncertainty in the long-term effects of the current motions of the robots by generating ‘snapshots’ of potential future scenarios. Specifically, each robot simulates possible plans of actions for a given time horizon, and after assessing each one of these simulated plans, it evaluates in ‘hindsight’ the quality of the first action of the plan. Each plan consists of a sequence of motion primitives. By performing multiple simulations with each initial action, the robot reduces the uncertainty associated with the long-term consequences of each one of them. With PHOP, robots are able to predict regions in the environment that will introduced motion constraints, allowing them to act accordingly (for example, by completely avoiding paths going through that region). Results of comparing PHOP with ORCA indicate lower travel time for the robots using PHOP.

Both my proposed learning and planning approaches improve the quality of the global motion of the robots in a decentralized manner, when no communication is possible. Robots adapt online to large-scale complex multi robot environments using only observations of the dynamics of other robots.

3. FUTURE WORK

I have a number of ideas for the next stages of my thesis work, from increasing the efficiency of the proposed approaches (to enhance their real-time applicability) to enabling lifelong learning in the navigation of the robots.

Specifically, I am researching into ways of incorporating an even broader range of robot behaviors without involving excessive exploration, which is a drawback of the current approach. Combining the motion simulation of PHOP with ALAN’s reward function might help in this research line, by enabling robots to simulate a range of short-term motions and execute the best evaluated one. Another plan is to increase the awareness of each robot with respect to the robots in its neighborhood. Instead of considering other robots only as part of the environment, I plan to use their individual motion patterns to learn context information. Specifically, their individual positions and velocities can provide useful information about future constraints in the environment before they directly affect the robot. With this information, robots might prefer to move together, forming groups in an emergent manner, if this is predicted to increase the efficiency of their collective motion.

Furthermore, I would like to extend the learning framework to enable robots to be continuously learning and reusing learned policies in new environments, empowering them to improve their motion behaviors in a lifelong learning process [5]. Several challenges need to be addressed to achieve this, such as how to transfer learned policies, generalize the learned knowledge, and many others.

Finally, I am also interested in environments where robots coexist and work in harmony with humans. To achieve this, robots should be able to navigate through the world in a human-like manner, and adapt to social conventions by learning and applying human-like actions.

REFERENCES

[1] E. Chong et al. A framework for simulation-based network control via Hindsight optimization. In *IEEE Conf. on Decision and Control*, volume 2, pages 1433–1438, 2000.

[2] J. Godoy, I. Karamouzas, S. J. Guy, and M. Gini. Anytime navigation with progressive hindsight optimization. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 730–735, 2014.

[3] J. Godoy, I. Karamouzas, S. J. Guy, and M. Gini. ϵ-UCB for action selection in multi agent navigation. In *Machine Learning and Planning for Control of Robot Motion Workshop at IROS*, 2014.

[4] J. Godoy, I. Karamouzas, S. J. Guy, and M. Gini. Adaptive learning for multi agent navigation. In *Proceedings of AAMAS*, 2015.

[5] S. Thrun and T. M. Mitchell. *Lifelong robot learning.* Springer, 1995.

[6] J. van den Berg, S. J. Guy, M. Lin, and D. Manocha. Reciprocal n-body collision avoidance. In *Robotics Research: The 14th International Symposium ISRR*, volume 70 of *Springer Tracts in Advanced Robotics*, pages 3–19. Springer-Verlag, 2011.