Learning To Move - Reinforcement Learning in Navigation

Brandon Haworth
University of Victoria
bhaworth@uvic.ca

1 Introduction

The simulation of human movement has become central to several areas of research and industry including animation, film, environment design, and emergency preparedness. These areas rely on methods that may be derived from several different principles or approaches. Each of these approaches embodies some assumptions about how humans are represented and how they move as a formalized model. Generally speaking, the movement models are comprised of several interacting sub-models that handle different conceptual layers of the navigation problem from behaviour to steering. Of particular interest are agent-based models where individual entities handle their decentralized actions and interactions. Macroscopic behaviours may then emerge from the microscopic agent-based model. The history of these models is rich and varied, starting with the need to animate numerous interacting characters in computer graphics [Reynolds, 1987]. Since then a plethora of models have emerged [Pelechano *et al.*, 2008; Thalmann and Musse, 2013; Huerre *et al.*,].

Of particular importance in the human simulation paradigm, crowd simulation, or synthetic crowds, is the lowest level model, i.e. steering and collision avoidance. These models handle the moment-to-moment movement decisions which resolve agent-agent and agent-environment collisions and goal-reaching behaviours. At this level, the agent often has a computationally efficient and largely simplified point-mass, or particle, representation. This improves the efficiency of collision detection and avoidance decisions while reducing the overhead required to store agent configurations, shapes, or plans. These affordances and efficiencies have made the single-particle agent-based model ubiquitous in research and industrial applications of synthetic crowds.

However recent work has shown that there are two key limitations in this approach. The first is that the underlying representation is not performant in terms of the representation of diverse humans and mobilities. Second, the production of these models often relies directly on encoded rules, expert decisions and beliefs, or fixed/limited datasets[Haworth, 2019]. This work focuses on exemplifying these issues through simulation-based analysis. The work first presents a large scale comparative quantitative and qualitative analysis of normalized rule-based single-particle steering models in contrast with spacetime planning multi-particle biomechanical steering model. This analysis highlights the contrast between model foundations and representations, showing that while all models have edge cases where performance is poor or collapse completely into fail cases, that biomechanically founded models afford a higher-fidelity and more human-like outcomes. Additionally, this work shows that representation and diversity in the steering model at the single agent level lead to significantly different outcomes at the crowd level (e.g. diversity impacts outcomes in safety-critical scenario simulation). Specifically, the work shows that the prior method of using desired agent velocity as a proxy for crowd heterogeneity (e.g. older agents are slower) is too limited in its ability to accurately represent human diversity in synthetic crowds. These results highlight the need for higher fidelity representations and for learning-based approaches for movement policies in steering models intended to represent humans. Here we present potential new solutions in the form of a groundbreaking synthesis of research areas in machine learning and physical character animation.

2 Learning to Move

Past methods for synthetics crowds can broadly be placed into two groups: data-driven and rule-based. Data-driven methods are limited because they tend to localize to the data. Rule-based methods are limited because they tend to meet the rules and can not account for all situations. Both approaches are highly dependent on their underlying representation. Specifically, single-particle representations of humans oversimplify the actions space and collision corridor of individual agent's which has a notable impact on crowd level outcomes. Advances in reinforcement learning (RL) and specifically DeepRL (DRL) and Multi-agent RL (MARL) have the potential to help us learn policies from simulations without defining a concrete and limiting model or dataset. These advances also allow us to work with much more complicated representations, such as footsteps or full-body articulated agents, by learning interacting policies for locomotion. This is inspired by the human sensorimotor locomotion loop where supraspinal input produces behavioural locomotion decisions and interacts with the complex system of central-pattern generators, motor/sensory neurons, and functional morphology to produce locomotor movements [Tucker *et al.*, 2015]. To address these issues, we present two approaches both predicated on Deep Reinforcement Learning, (1) parametric policy learning for an agent-based single-particle steering model; and (2) hierarchical reinforcement learning for physical full-body humanoid crowds.

In this work, we first sought to apply this approach to recreating the prior art in the field, that is, to learn movement policies on top of the representations and heterogeneity of classic crowd simulators–the single-particle agent-based model with heterogeneous desired velocities and agent radii as a proxy for diverse human agents. By learning a single shared policy in a decentralized agent-based approach we can produce policies for agent movement which are akin to past crowd steering simulators. We found that useful steering policies can be learnt by using domain randomization and large scale simulation, i.e. giving agents random goals in randomized environments where their long term path is known. To learn diverse movements, however, presents a more complicated problem. Agents must 1) observe quantities from the environment which are not directly known to them; 2) be parametric, i.e. have a parameter that can be set by the practitioner after training has been completed and 3) learn policies that respond to other heterogeneous agents. Problems 1 & 3 are fundamentally related in reinforcement learning and be formalized under the concept of partial observability in the Markov decision process. For example, a person is not given the exact speed of a nearby person, instead, we infer this through sequential vision. We show that sequential observation stacking of the agent's vision (last N snapshots of the visual field, in this case, depth rays) allows agents to learn velocity and acceleration based policies without directly observing those quantities. We show that problem 2 can be solved by making the parameter ubiquitous in the reinforcement learning paradigm, i.e. observed in the state input (goal-conditioning or simply within observations); encouraged by the reward function during training; and utilized or consequential in the action space. By randomizing the parameter over the valid parameter range during training it becomes part of the domain randomization strategy. In this way, the agent learns a parametric policy or policy subspaces. That is, in a combinatorial manner, agents of a particular parameter setting learn to steer with agents of other parameter settings. To exemplify this method, we show that it works exceptionally well over the desired speed and agent radius parameters. Additionally, in many settings, the learned policy outperforms prior work by learning sequential policies that mimic higher-level longer-term planning, such as avoiding an area that is too densely occupied.

Our second approach seeks to bring together two disparate fields in robotics and animation to produce extremely high fidelity crowd steering simulators. Specifically, we marry crowd simulation and physical character control by proposing a hierarchical multi-agent reinforcement learning technique for agent-based physically interactive full-body humanoid crowds. The problem is structured as hierarchical by separating low-level control of the functional morphology, proprioception, and cyclic pattern generation from high-level control of environment observation, short- & long-term planning, and intelligence. Using similar domain randomization strategies as mentioned above we train the hierarchy in a bottom-up fashion. First, we learn a goal conditioned stable locomotion lower-level controller policy that produces torques for joint level PD controllers in the character. This lower-level policy is conditioned on a two-step foot placement plan which the higher-level policy is meant to produce. During training, this two-step plan is sampled across a broad spectrum of difficulties and the character undergoes external randomized forces that mimic those that humans may encounter during dense or adverse crowd scenarios. The resultant low-level policy is shared among agents in a fashion similar to a common movement language–this ideally reduces the impact of the non-stationarity problem encountered in MARL. The high-level policy takes in environment observations including an egocentric relative velocity field. We train the high-level policy using domain randomization as above. We show that using only simple navigation goal-reaching rewards the high-level policy learns complex navigation strategies. We also show that the high-level policy can be heterogeneous and rewards may be substituted or combined to produce arbitrary behaviours such as navigation, tag games, soccer games, etc[Haworth *et al.*, 2020; Berseth *et al.*, 2019].

References

[Berseth *et al.*, 2019] Glen Berseth, Seonghyeon Moon, Mubbasir Kapadia, Petros Faloutsos, et al. Multi-agent hierarchical reinforcement learning for humanoid navigation. In *Deep Reinforcement Learning Worskop at NeurIPS 2019*, 2019.

[Haworth *et al.*, 2020] Brandon Haworth, Glen Berseth, Seonghyeon Moon, Petros Faloutsos, and Mubbasir Kapadia. Deep integration of physical humanoid control and crowd navigation. In *Motion, Interaction and Games*, pages 1–10. 2020.

[Haworth, 2019] M. Brandon Haworth. *Biomechanical Locomotion Heterogeneity in Synthetic Crowds*. PhD thesis, York University, Toronto, Canada, November 2019.

[Huerre *et al.*,] Stephanie Huerre, Jehee Lee, Ming Lin, and Carol O'Sullivan. Simulating believable crowd and group behaviors. In *ACM SIGGRAPH ASIA 2010 Courses*, pages 13:1–13:92.

[Pelechano *et al.*, 2008] Nuria Pelechano, Jan M. Allbeck, and Norman I. Badler. *Virtual Crowds: Methods, Simulation, and Control*. Morgan & Claypool Publishers, 2008.

[Reynolds, 1987] Craig W Reynolds. Flocks, herds and schools: A distributed behavioral model. In *ACM Siggraph Computer Graphics*, volume 21, pages 25–34. ACM, 1987.

[Thalmann and Musse, 2013] Daniel Thalmann and Soraia Raupp Musse. *Crowd Simulation, Second Edition*. Springer, 2013.

[Tucker *et al.*, 2015] Michael R Tucker, Jeremy Olivier, Anna Pagel, Hannes Bleuler, Mohamed Bouri, Olivier Lambercy, José del R Millán, Robert Riener, Heike Vallery, and Roger Gassert. Control strategies for active lower extremity prosthetics and orthotics: a review. *Journal of neuroengineering and rehabilitation*, 12(1):1, 2015.