SSM
Scuola Superiore Meridionale

MERC PhD Project Proposal 2023/2024

Title of the research project:

Machine-Learning based control of autonomous multiagent systems for search-and-rescue operations in natural disasters

Keywords (up to five)

Autonomous systems; complex systems; control theory; game theory; machine learning

Supervisors (at least two from two different areas):

Supervisor 1 (name, contact details, homepage, area of expertise)
Mario di Bernardo, MERC, complex systems, control theory

Supervisor 2 (name, contact details, homepage, area of expertise)
Mirco Musolesi, MERC, machine learning. Reinforcement learning, computer science

Supervisor 3 (name, contact details, homepage, area of expertise)
Giovanni Russo, MERC, control theory, Reinforcement learning

Supervisor 4 (name, contact details, homepage, area of expertise)
Michael Richardson, MERC, herding, human behaviour, modeling

Project description (max 5000 characters)

Creating effective distributed control strategies for large-scale multi-agent systems, such as in swarm robotics, remains a critical challenge. Traditional centralised approaches lack robustness to disruptions and scale poorly. This project aims at developing strategies allowing technology-based multi-agent systems to learn coordination and cooperation through distributed decision-making, inspired by natural systems' adaptability. With the limitations of traditional control theory in handling the complexity of realistic multi-agent systems, machine learning, especially multi-agent reinforcement learning (MARL), has emerged as a promising solution. MARL, which models each agent's decisions and interactions as a Markov decision process, has shown success in various applications, including robotics and smart grids. However, it faces significant challenges like the curse of dimensionality and the need for improved data efficiency as the number of agents grows.

To tackle the challenges in multi-agent reinforcement learning (MARL), literature suggests strategies that integrate models or physical principles to inform the learning process, utilizing system knowledge to shorten learning times, require less data, and ensure stability and safety. Another method combines learning with control strategies, such as integrating reinforcement learning (RL) with model predictive control (MPC), to enhance learning efficiency and control performance. A "control-tutored" approach, where a learning agent is guided by a control strategy with limited system knowledge, has proven effective in addressing testbed challenges from Open AI Gym.

Exploring the integration of control strategies with multi-agent reinforcement learning (MARL) for multi-agent systems, especially in search and rescue (SAR) operations post-natural disasters, offers significant potential. This approach remains largely unexplored and could crucially enhance the coordination of autonomous agents (like drones) in executing collective tasks within unpredictable environments. SAR operations, occurring in diverse settings such as wilderness, sea, or urban areas, present unique challenges including task allocation, path planning, area exploration, and rescuee transportation. While MARL solutions for SAR have been proposed, their practical application is limited by oversimplified assumptions, highlighting the need for more adaptable and realistic approaches.

The *objectives* of this project can be listed as follows:

1) Development of efficient learning-based control strategies for complex multi-agent systems.

   This objective is aimed at synthesising innovative strategies for controlling the collective behaviour of cooperative multi-agent systems performing a joint desired task by combining MARL and control theoretic approaches to steer the dynamics of complex systems. This will be instrumental for the development of new techniques to solve SAR problems.

2) Analysis of the convergence and robustness of the developed methods

   This objective concerns the investigation of the convergence and robustness properties of each of the strategies developed within the project on a set of multi-agent benchmarks problems. This will allow for the comparison of the methodologies with those already existing in the literature, which will be studied as part of [O1].

3) Apply the strategies to search and rescue problems in natural disasters where autonomous agents are tasked with the goal of corralling and bring to safety another group of agents.

   This will entail deploying the strategies on more realistic models of agents' behaviour that might include the case of evading target agents in a multi-agent pursuit-evasion scenario.

Methodology

To achieve [O1], we will start from the control-tutored (CT) approach to learning we recently presented, where, to achieve the control of a single system of interest, a tabular learning strategy (e.g. Q-Learning) is combined with a state feedback controller to achieve better learning and control performance. We named the resulting strategy as "Control-Tutored Reinforcement Learning" (CTRL). Therein, the controller is designed assuming limited knowledge of the uncertain system of interest and a policy is proposed that, according to some criterion, selects at each step either the action suggested by the tabular strategy or that suggested by the controller. In particular, we explored both the use of a deterministic criterion and of a probabilistic one to inform the choice of the best action to take. In the deterministic setting, at each step the agent is presented with two actions (one from the tabular strategy and one from the controller) and uses the one that maximises or minimises an appropriately chosen cost function related with the control goal to be achieved. In the stochastic setting, instead, at each step the policy is selected by randomly choosing one of the two actions with a given probability

A first extension of CTRL we intend to carry out in this project is to combine Deep Reinforcement Learning (DRL) algorithms or policy-based algorithm rather than tabular strategies with control tutors (developing a control-tutored DRL or CT-DRL approach) in order to achieve an even better performance, particularly in the presence of larger uncertainties, continuous state/action spaces and mismatches between the real system and the model used to devise the control law to be combined with the learning algorithm.

The next step will be that of designing control tutored MARL strategies (CT-MARL). This will be achieved by combining existing MARL policies with distributed control strategies for network systems that have been recently presented in the literature to steer their collective behaviour. In particular, we will start with the problem of achieving convergence and synchronization in complex multi-agent systems, focussing on leader-follower and consensus problems as a testbed example. In this context, agents will need to learn how to coordinate their behaviour and cooperate with each other in order to achieve synchronous behaviour on the basis of information coming only from those neighbouring agents within the ensemble they are connected with. In the spirit of CTRL, we will construct policies where each agent can select its next action either by using the policy suggested by a CT-DRL algorithm running locally or that proposed by a network control system based on a limited model of the agents behaviour (typically encoded in a model of the ensemble as a network of interacting nonlinear dynamical systems).

To fully leverage the benefits of CTRL strategies we will also explore the idea of using Hierarchical Multi Agent Reinforcement Learning (HMARL), where we will decompose the overall task into a set of subtasks that the agent will have to master using control-tutored strategies. Decomposing the control objective in subtasks has several advantages, such as reducing the impact of the curse of dimensionality happening with large state spaces, as is the case with multi-agent problems. Additionally, this framework will allow us to designing ad-hoc control laws to tutor the learning agents in each of the subtasks.

We will explore both the use of deterministic and stochastic criteria in informing the policies that agents can use to decide their next action among those suggested by the RL or CT algorithms. We will also study the case in which different agents use different criteria to select their policies, investigating for the first time such a heterogeneous scenario. The latter is particularly apt to orchestrate the behaviour of agents of different types (such as, for example, those being used for certain SAR operations involving the concurrent use of different types of ground and aerial vehicles).

We will achieve [O2] by carrying out the analysis of all the strategies developed in the project, with the aim of (i) assessing their convergence and robustness properties, and (ii) establishing what properties the learning agents inherited from those encoded in the tutoring control algorithms used to inform the learning. We will start with an extensive numerical analysis of their properties by using a set of representative multi-agent testbed applications, which will be used as benchmarks. A possibility will be also to explore the use of density control strategies where the dynamics of the agents involved is approximated at the macroscopic level via PDEs.

Finalle to achieve [O3] we will consider a more realistic simulation scenario. In this context we will have to account for possible obstacles in the arena where the agents move, possibly differences in the terrain. This scenario will also entail the problem of searching and navigate the environment

to locate targets to rescue to move them from unsafe to safe areas. The target dynamics to be considered will include the case of random target dynamics, flocking targets and/or evading targets.

Relevance to the MERC PhD Program (max 2000 characters)

The control of multiagent systems is central to the focus of the MERC PhD program as it entails finding strategies to control large scale complex systems. Its numerous applications spanning from multi-robots systems to the study of animal groups is also of great relevance in the areas of interest to the PhD program. The work will be carried out within the scope of the National PRIN 2022 Project that involves both prof di Bernardo (as national coordinator) and prof Musolesi (as team-leader at the University of Bologna and UCL).

Key references

F. de Lellis, M. Coraggio, **G. Russo**, M. Musolesi, M. di Bernardo, "CT-DQN: Control-Tutored Deep Reinforcement Learning", 5th Annual Learning for Dynamics & Control, 2023

F. de Lellis, **G. Russo**, M. di Bernardo, "Tutoring Reinforcement Learning via Feedback Control", 2021 European Control Conference

M. Rathi, P. Ferraro, **G. Russo** "Driving Reinforcement Learning with Models", Intellisys 2020, September 2-3, 2020

Pierson, A. and Schwager, M., 2015, May. Bio-inspired non-cooperative multi-robot herding. In *2015 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1843-1849). IEEE.

Scott, W. and Leonard, N.E., 2013, June. Pursuit, herding and evasion: A three-agent model of caribou predation. In *2013 American Control Conference* (pp. 2978-2983). IEEE.

Auletta, F., Fiore, D., Richardson, M.J. and di Bernardo, M., 2022. Herding stochastic autonomous agents via local control rules and online target selection strategies. *Autonomous Robots*, *46*(3), pp.469-481.

Nalepka, P., Lamb, M., Kallen, R.W., Shockley, K., Chemero, A., Saltzman, E. and Richardson, M.J., 2019. Human social motor solutions for human–machine interaction in dynamical task contexts. *Proceedings of the National Academy of Sciences*, *116*(4), pp.1437-1446.

Nikitin, D., Canudas-De-Wit, C. and Frasca, P., 2021. Boundary control for stabilization of large-scale networks through the continuation method.

Escobedo, R., Ibañez, A. and Zuazua, E., 2016. Optimal strategies for driving a mobile agent in a “guidance by repulsion” model. *Communications in Nonlinear Science and Numerical Simulation*, *39*, pp.58-72

Joint supervision arrangements

The supervisors will work closely with the PhD student for the successful completion of the PhD thesis. The student will meet with at least one of the supervisors on a weekly basis while there will be other regular meetings with all supervisors, including during the period abroad when such meetings will be held remotely.

Location and length of the study period abroad (min 12 months)

The student will spend periods abroad working in the group of prof Mirco Musolesi at UCL (UK) and prof Mike Richardson (Sydney)

Any other useful information

Further shorter research visit abroad will be planned to work in collaboration with experts in game theory and machine learning and attend PhD courses and relevant workshops.