SSM
Scuola Superiore Meridionale

MERC PhD Project Proposal 2023/2024

Title of the research project:

Towards Autonomous Agents That Learn, Control and Optimize Like People

Keywords (up to five)

Autonomous Agents, (Multi-Agent) Reinforcement Learning, Data-Driven Control, Optimal Control

Supervisors (at least two from two different areas):

Supervisor 1: *Giovanni Russo (giovarusso@unisa.it)*
Dept. of Information and Electrical Engineering & Applied Mathematics (DIEM) at University of Salerno
www.sites.google.com/view/giovanni-russo
Reinforcement Learning, Data-Driven Control, Control Theory, Complex Systems, Optimization

Supervisor 2: *Mario di Bernardo (mario.dibernardo@unina.it)*
Department of Electrical Engineering and Information Technology, University of Naples Federico II and Scuola Superiore Meridionale
https://sites.google.com/site/dibernardogroup/home
Complex Systems, Learning, Nonlinear Dynamics

Project description (max 5000 characters)

The vision and the context of the project. Despite recent achievements, we are better than machines at solving a range of difficult problems: we re-use what we learned, abstracting and adapting the knowledge we gained to solve new tasks in new situations. Machines lack these abilities, and this is in fact the biggest challenge towards real-world deployment of autonomous agents that learn efficiently and reliably in different operating conditions, across different tasks. In this context, the **overarching goal** of this project is that of designing novel decision-making mechanisms enabling agents to emulate our ability of learning new tasks by re-using and abstracting knowledge, while iteratively improving performance without any re-training.

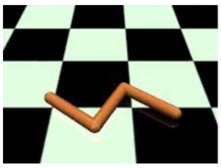
Gathering inspiration from neuroscience and cognitive sciences, we will achieve this goal by pursuing a rigorous approach (see also below) grounding the design of autonomous agents on the Nobel-winning fast/slow thinking theory (FST) from cognitive sciences and on the thousands brain theory (TBT). The **results** will be relevant to all applications, transversal to learning and control, where agents need to reliably fulfil tasks requiring physical interactions with unknown, non-stationary and stochastic environments, such as the *open sea* in the proposed application, and where there is not enough data to devise first-principles models.

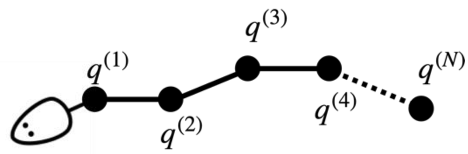
Project objectives and methodology. The detailed research program will be shaped based on the **interests of students**. A preliminary list of concrete research objectives is given below. The objectives are **inherently ambitious** and PhD students, also guided by their supervisors, will have the opportunity to selectively focus on a subset of specific objectives, in accordance with their individual interests, available time and opportunities.

O1 – *Towards a Thousand Brain Theory For Control and Learning***.** It is widely believed that our ability to grasp new concepts by re-using knowledge is hosted in our neocortex, a highly redundant distributed system; its units are cortical columns, specialized to process data. Each column is designed to quickly output a decision based on the data it sees and TBT then postulates that the policies determining our behavior arise from composing the decisions from the columns. Moreover, with thousands of units processing the stream of data, the distributed nature of the neocortex guarantees that our decisions are resilient towards faults of (most) individual columns. With this objective, inspired by some prior work from one of the proponents, we aim to mirror this fundamental mechanism. We will do so by designing a neocortical-inspired algorithm that, based on the available data, composes the outputs from a network of units providing primitive policies tackling elementary tasks. For concreteness, we will early benchmark the results on a simulated use-case where a swimming agent needs move within an unknown environment (see figure below). In this context, the primitive policies will be elementary motion primitives (e.g., moving one joint of the agent) and the neocortical algorithm will need to properly patch the primitives to allow proper coordinated movements of the joints, so that the agent can effectively swim.

O2 – Embedding Fast/Slow Thinking in Control and Learning. The FST argues that we can efficiently compute actions thanks to a dichotomy between two systems in our brain; the fast system (FS) designed to operate with little or no effort, and the slow system (SS) allocating attention to effortful mental activities. We use FS most of the times as it can efficiently generate patterns of decisions based on heuristics and cognitive biases. The decisions from FS are parsed by SS, which, if not satisfied with FS computations, engages additional resources to make more informed decisions. In this spirit, the algorithm from O1 belongs to the FS and, within this objective, we will design an offline algorithm that will have the goal of refining the policy by learning a model of the environment and a *better* cost to execute the task. These refined models/costs will be used to compute a brand new policy that will both bias the actions of the agent the next time the task is repeated and enable flexibility to tackle new situations. Based on some of the preliminary studies from one of the proponents, the mechanism to tackle O2 might leverage Inverse Reinforcement Learning via convex optimization. Again, we will benchmark the results by considering the use case from O1, this time with the agent moving in an environment filled with objects and other swimmers. We will then verify that the results of this objective enable the agent to tackle this more challenging environment without or little training.

O3 – Architecture Embodiment. We will ultimately embody the results from the above objectives. While the application can be tailored towards the specific interests of the students, we plan to validate the results on the use-case from O1 and O2, this time using the policies to enable a team of autonomous robotic swimmers to cooperatively locate and collect objects (e.g., pollutants). The policies will be deployed on e.g., the simulated C. elegans-like robot from MuJoCo (this is an articulated body, as shown in the figure). Depending on time and opportunity, we will explore the possibility of deploying the algorithms onto a real robotic agent, such a rover.





Workplan. The project offers a mix of theory and experimentations and the workplan, tailored towards students' background, will be built to reflect this aspect giving to the project either a more methodological- or application-oriented angle: given the high ambition of each individual objective, students will have the opportunity to selectively focus on a subset of the objectives. The plan will be developed in incremental tasks and periodic meetings will be scheduled with the supervisors. First, the student will start with becoming familiar with the existing literature in the areas related to the project. The output of this first step will be the definition of a preliminary methodology and a refined application. Then, in the second phase, the student will develop the methodology and perform deployments/tests on small-scale problems (as described in the project objectives). The final part of the project will see the student finalizing the methods and tools deploying on the selected application. Simulated test-cases are available to students: these already available cases can be leveraged as an *entry point* with the goal of facilitating students towards familiarizing with both the application and the methodological concepts that will be developed. Within the project, students will also benefit from visiting labs at the forefront of the topics covered in the above objectives (see below for potential destinations/collaborations).

See the list of references for further details on the different aspects of the project.

Relevance to the MERC PhD Program (max 2000 characters)

This **ambitious** project is, by its very definition, **interdisciplinary**. We aim at designing *better* autonomous agents contaminating control, learning and optimization with ideas from other fields as outlined above. Depending on the students' interests, the main outcomes of the project will be a combination of novel methodological and application-oriented results. From the methodological viewpoint, the main outcome of the project will be a novel methodology enabling effective and efficient learning of new tasks from data. From the applications viewpoint, the results might lead to a novel, and resilient, architecture for efficient decision-making. The methodology and the architecture will allow to tackle an application (as described above) that is difficult, if not impossible with classical control and learning methods. The link with complexity and risk is apparent: (i) risk elements will be embedded by considering tasks for which agents need to make optimal decisions to safely optimize a risk index; (ii) we expect that the decision-making architecture itself will be a complex network, for which the key emerging characteristics will need to be analyzed.

See the list of references for further details on the different aspects of the project.

Key references

A. Tversky, D. Kahneman "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty," Science, vol. 185, pp. 1124-1131, 1974.

J. Hawkins, A thousand brains: a new theory of intelligence, Basic Books, 2021.

D. Kahneman, Thinking, fast and slow, Penguin Books, 2011.

V.B. Mountcastle, "An organizing principle for cerebral function: the unit module and the distributed system," The mindful brain, 1978.

V.B. Mountcastle, B.C. Motter, M.A. Steinmetz, A.K. Sestokas, "Common and differential effects of attentive fixation on the excitability of parietal and prestriate cortical visual neurons in the macaque monkey," Journal of Neuroscience, vol. 7, pp. 2239-2255, 1987.

E. Garrabé, G. Russo, Probabilistic design of optimal sequential decision-making algorithms in learning and control, Annual Reviews in Control, vol. 54, pp. 81-102, 2022

D. Gagliardi, G. Russo, "On a probabilistic approach to synthesize control policies from example datasets", Automatica, vol. 137, 110121, 2022

E. Garrabé, G. Russo, "On the design of autonomous agents from multiple data sources", IEEE Control Systems Letters, vol. 6, pp. 698 - 703, 2022

G. Russo, "On the crowdsourcing of behaviors for autonomous agents", IEEE Control Systems Letters, vol. 5, pp. 1321-1326, 2021

B. M. Lake et al., "Human-level concept learning through probabilistic program induction," Science, vol. 350, no. 6266, pp. 1332–1338, 2015

E. Garrabé, G. Russo, "CRAWLING: a Crowdsourcing Algorithm on Wheels for Smart Parking", Scientific Reports, vol. 13, 16617, 2023 Collection in Advanced Transportation Systems

E. Garrabé, M. Lamberti, G. Russo, "Optimal Decision-Making for Autonomous Agents via Data Composition", IEEE Control Systems Letters, vol. 7, pp. 2557-2562, 2023 – also presented at the IEEE 62nd Conference on Decision and Control, CDC 2023

Karl Tuyls and Gerhard Weiss. Multiagent Learning: Basics, Challenges and Prospects. AI Magazine. Volume 33. Issue 3. 2012.

R. M. Murray, K. J. Astrom, S. P. Boyd, R. W. Brockett, G. Stein. Future directions in control in an information-rich world. IEEE Control Systems Magazine. Volume 23. March 2003.

Joint supervision arrangements

Meetings will be scheduled on an as-needed basis, in order to ensure the effective development of the project. As a minimum, supervisor(s) will meet students at least once a week.

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

The candidate will be able to spend a research period (or research periods) at one or more labs at the forefront of research in the topics that are relevant for the proposed research plan. Specifically, within this project, students will have the opportunity to spend time abroad at one or more of the following potential destinations:

(i) the group of Prof. Florian Dörfler at ETH Zurich;

(ii) the group of Professor Robert Shorten at the Dyson School of Design Engineering within Imperial College London.

Any other useful information

While the project can be tailored towards students with more application-oriented interests, the project is best suited for students with a preference towards mathematical rigour (a proof-oriented mindset) and with a background in dynamical systems, control and learning. For further details on the background students can contact their potential supervisors.

Please return this form via email by no later than 9th February 2024 to merc@ssmeridionale.it