Thesis Proposals
Machine Learning

Prof. Marcello Restelli
February 2022

Machine Learning Team

Marcello Restelli
Team Leader

Francesco Trovò
Team Leader

Alberto Metelli
PostDoc Researcher

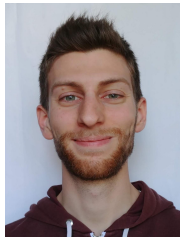
Giorgio Manganini
Assistant Professor

Lorenzo Bisi
PhD Student

Edoardo Vittori
PhD Student

Mirco Mutti
PhD Student

Alessandro Nuara
CTO @ MLCube



Alessandro Lavelli
ML Eng @ MLCube

Diego Piccinotti
ML Eng @ MLCube

Amarildo Likmeta
PhD Student

Pierre Liotet
PhD Student

Marco Mussi
PhD Student

Luca Sabbioni
PhD Student

Alessio Russo
PhD Student

Gianluca Drappo
PhD Student

Riccardo Zamboni
Research Scholar

Riccardo Poiani
PhD Student

Paolo Bonetti
PhD Student

Outline

- Three types of theses
  - Industrial theses
  - Internship theses
  - Research theses
- Application instruction

Industrial Theses

IMICIB - Intesa Sanpaolo

RL for Market Making

Bid		Offer	
Volume	Price	Price	Volume
3100	2.85	3.00	1000
4900	2.80	3.05	4800
		3.10	2200

Market makers provide liquidity to the markets by continuously pricing an asset in bid and offer

Goal: Develop a RL market making agent

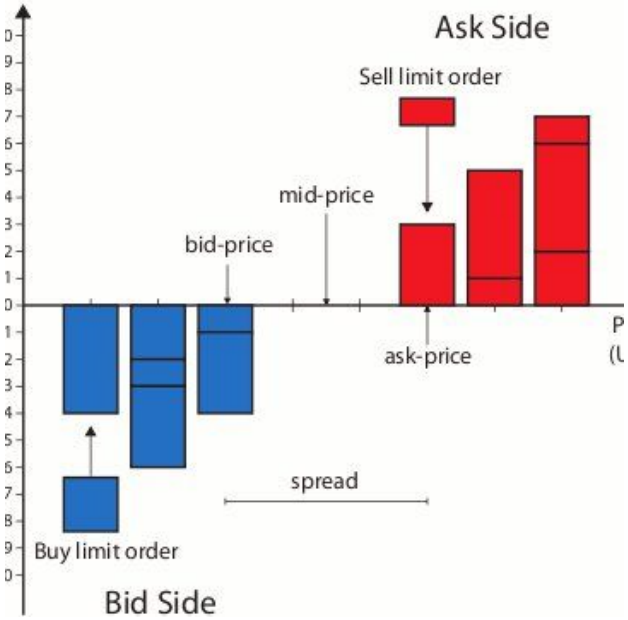
Steps:

- Reproduce the state of the art
- Learn in a market simulator
- Test in crypto market?

Supervised by: Edoardo, Lorenzo, Luca, Pierre
Timespan: 9-12 months
Start: immediately
Contact: edoardo.vittori@polimi.it

IMICIB - Intesa Sanpaolo

RL for Trading - Towards more realistic agents



Simple financial MDPs allows to only buy or sell an asset at the **current best price** (*market order*).

We want to consider agents that can:

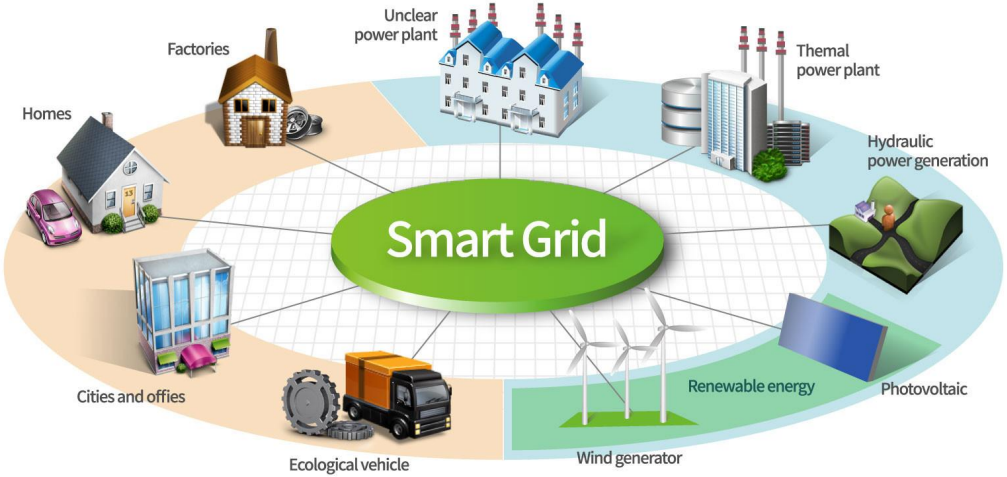
- emit **limit orders**
- modify **previous orders**

Goal: Develop a more realistic RL agent, which can act by means of limit orders.

Supervised by: Lorenzo (with Luca, Pierre and Edoardo)
Timespan: 9-12 months
Start: immediately
Contact: lorenzo.bisi@polimi.it

RSE - Ricerca Sistema Energetico

Management of energy storage systems in Smart Grids



Crucial issue about the degradation of the storage systems (battery) due to usage

Currently no or few policies have been adopted to preserve and manage storage system

Objective: use of RL techniques to design policies to handle the degradation in an optimized way (constrained by the battery usage)

Supervised by: Francesco
Timespan: 9-12 months
Start: a few months
Contact: francesco1.trovo@polimi.it

SIEMENS

MARL for Industry 4.0

Provided a digital twin of a factory/production floor, we want to **develop effective MARL algorithms** for solving a problem of **cooperation** with **partial observability**, **decentralization** and **potentially multiple tasks.**

Goal: Develop and adapt state-of-the-art (SOTA) MARL algorithms to our specific problem

Steps:

- Study, understand and select SOTA algorithms
- Implement them in a (almost developed) simulated environment

Requirements: Good programming skills in Python (and possibly some C++)

Supervised by: Riccardo Z.
Timespan: 9-12 Months
Start: April-May 2022
Contact: `riccardo.zamboni@polimi.it`

LEONARDO

Multi Agent Hierarchical RL for autonomous aircraft

Supervised by: Gianluca
Timespan: 9-12 months
Start: March-April 2022
Contact: gianluca.drappo@polimi.it

Setting

Autonomous management of heterogeneous team of aircraft during mission execution in contested operating environment using **Multi-Agent Hierarchical Reinforcement Learning** approaches

Objectives:

- Mission elaboration → hierarchical policy
- Replanning for unexpected behaviour

Internship Theses

Yeldo

Machine Learning for real estate investments

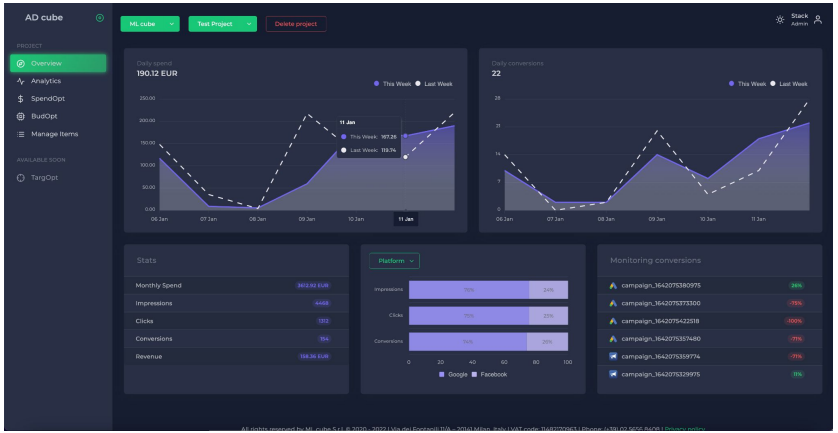
- **Goal**: Automatic evaluation of real estate investments opportunities
- **Available Data**:
  - Historical data of previous investments
  - Property features (e.g., position, points of interest, usage, connections, etc.)
- **Requirements:**
  - Good programming skills
  - Teamwork skills

Supervised by: Francesco Trovò and Alessandro Nuara
Timespan: 9-12 months
Start: March
Contact: alessandro.nuara@mlcube.com

ML cube

AD cube

Online Machine Learning Algorithms for Advertising Campaigns Optimization



- **Goal**: implement and deploy online machine learning algorithms for advertising campaigns optimization
- **Activity**: work with the ML cube team on the development of AD cube
- **Topics**:
  - safe budget optimization,
  - model selection
  - MLops
- **Requirements:**
  - Good programming skills in Python
  - Teamwork skills
  - Plus: experience with AWS services and DevOps

Supervised by: Alessandro Nuara
Timespan: 9-12 months
Start: Now
Contact: alessandro.nuara@mlcube.com

ML cube

AD cube

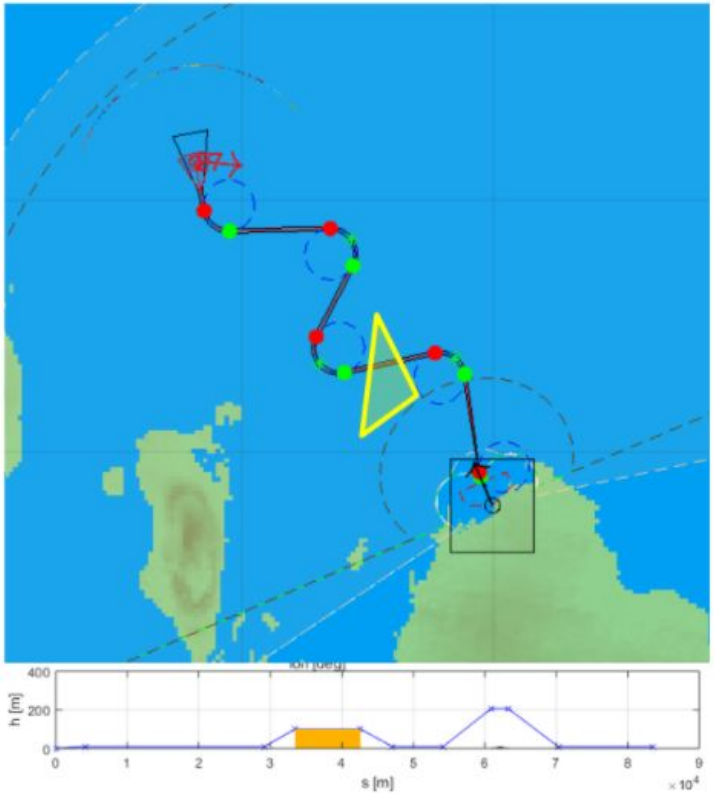
Machine Learning Algorithms for Targeting Optimization

- **Goal**: Identify the optimal target of digital advertising campaigns
- **Data Sources:**
  - Google Ads and Facebook Ads API
  - Web-site traffic data
  - Scraping
- **Requirements:**
  - Good Python programming skills
  - Teamwork skills

Supervised by: Alessandro Nuara and Diego Piccinotti
Timespan: 9-12 months
Start: Now
Contact: alessandro.nuara@mlcube.com

ML cube

MBDA - Mission Planning



Goals:

- Missile optimal **trajectory** generation, based on several criteria
- **Real-time** requirements: fast update
- **Coordination** between different weapons
- POC in **Matlab**

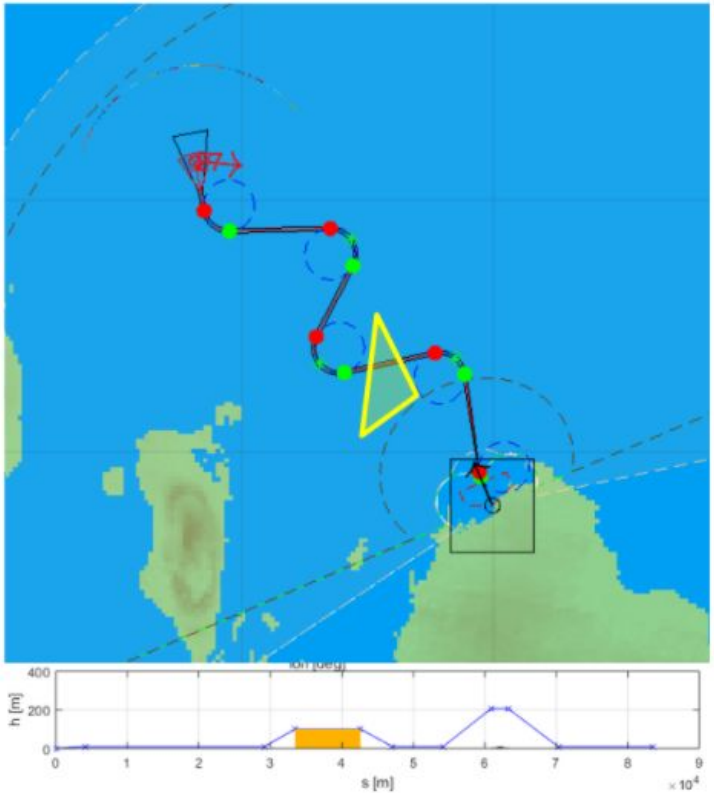
Methodologies: automatic **planning**

- Multi-objective
- Any-time

Supervised by: Alberto Metelli, Alberto Marchesi, MLCube team
Timespan: 12 months
Start: April
Contact: albertomaria.metelli@polimi.it

ML cube

MBDA - Performance Modeling



Goals:

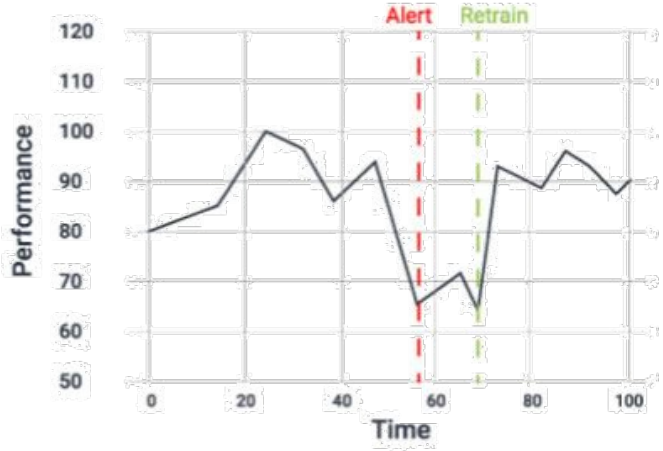
- Given a **kinematic** trajectory, we want to estimate
  - Flight time
  - Fuel consumption
  - Minimum speed along the trajectory
  - Feasibility
- **Point-wise** estimation of consumption, times and speeds

Methodology: uncertainty-aware estimation

Supervised by: Alberto Metelli, Alberto Marchesi, MLCube team
Timespan: 12 months
Start: April
Contact: albertomaria.metelli@polimi.it

ML cube

ML cube Platform
Monitoring & Automatic Retrain of ML Models



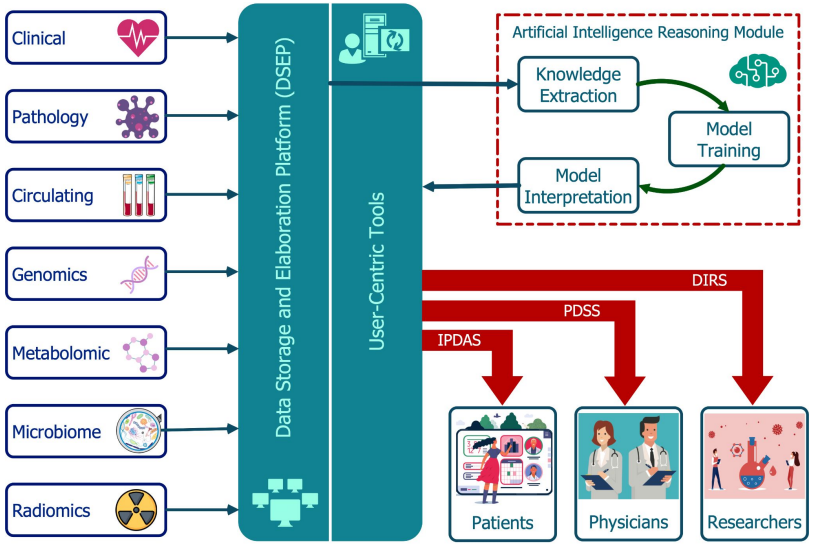
- **Goal**: implement algorithms for monitoring and automatic retrain of ML models
- **Activity**: work with the ML cube team on the development of the ML cube Platform 1st release
- **Requirements:**
  - Good programming skills in Python
  - Teamwork skills
  - Plus: Experience with DevOps and MLOps technologies

Supervised by: Alessio Russo and MLcube Team
Timespan: 9-12 months
Start: March
Contact: `alessio.russo@polimi.it`

ML cube

Integrative science, Intelligent data platform for Individualized LUNG cancer care with Immunotherapy

I³ LUNG



- **Goal**: Implement a platform to integrate the expertise and data provided by different institutions (e.g., Istituto Nazionale dei Tumori, Istituto Europeo di Oncologia, Lung Cancer Europe) to provide the access to a ML-based predictive tool for immunotherapy treatments

- **Requirements**:
  - Good Programming and Software development skills

Supervised by: Alessandro Nuara and Francesco Trovò
Timespan: 9 months
Start: March
Contact: `alessandro.nuara@mlcube.com`

ML cube

Dynamic Pricing for Hotels

- **Goal**: implement algorithms for dynamic pricing of hotels' rooms
- **Data Sources**: Competitors data, hotels historical data, demand, rooms availability, etc.
- **Requirements:**
  - Good programming skills in Python
  - Teamwork skills

Supervised by: Alessandro N.
Timespan: 9-12 Months
Start: a few months
Contact: alessandro.nuara@mlcube.com

ML cube

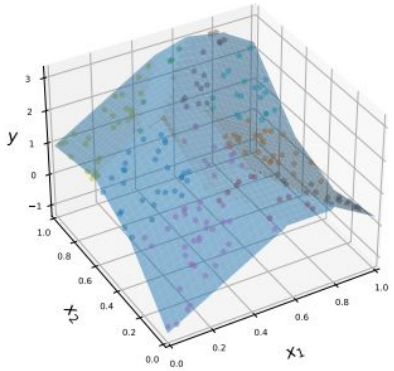
Dynamic Pricing for E-commerce

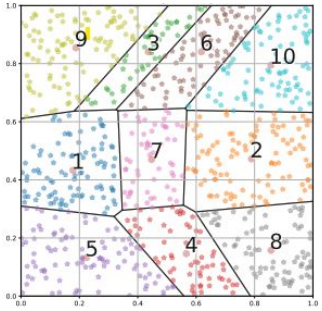
- **Goal**: Design and deploy algorithms for dynamic pricing of e-commerce products
- **Data Sources**: Competitors data, historical orders data, etc.
- **Requirements:**
  - Good programming skills in Python
  - Teamwork skills

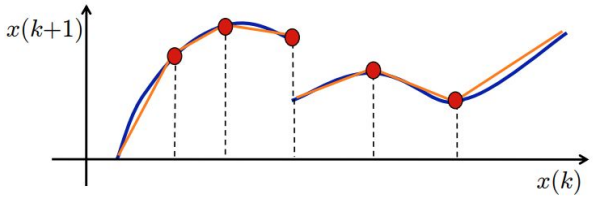
Supervised by: Alessandro N.
Timespan: 9 months
Start: March
Contact: `alessandro.nuara@mlcube.com`

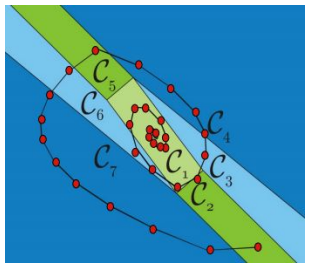
Research Theses

RL for Cyber-physical Systems Identification









WHAT: Multi-step Prediction Model Identification of PWA/PWARX models and dynamical systems from observed data

HOW: Reinforcement Learning and Evolutionary Algorithms as iterative optimization schemes

WHY:

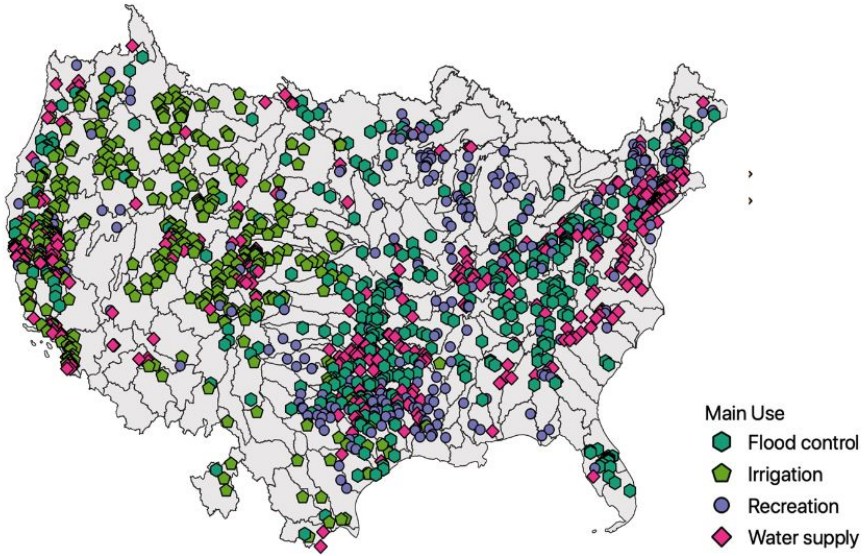
- Technological push to study cyber-physical models
- Central importance in a variety of model-based engineering applications
- Study and improvement of RL/Evolutionary algorithms

$$\begin{aligned} x(k+1) &= A_{i(k)}x(k) + B_{i(k)}u(k) + f_{i(k)} \\ y(k) &= C_{i(k)}x(k) + D_{i(k)}u(k) + g_{i(k)} \\ i(k) \quad &\text{s.t.} \quad H_{i(k)}x(k) + J_{i(k)}u(k) \leq K_{i(k)} \end{aligned}$$

$$f(x) = \begin{cases} F_1 x + g_1 & \text{if } H_1 x \leq K_1 \\ \vdots & \\ F_s x + g_s & \text{if } H_s x \leq K_s \end{cases}$$

Supervised: Giorgio
Expected time for graduation 9-12 months
Start: now
Contact: `giorgio.manganini@gssi.it`

Inverse Reinforcement Learning for Water Reservoir



Setting: dams impact on human systems: (i) reliable water supply (agriculture, power, and public supply), (ii) year-round navigation, (iii) support regional economic development, …

Goal: *Inverse Reinforcement Learning* => Infer the **intent** (aka **reward function**) of dam operators

ResOpsUS: historical inflows, outflows and changes in storage for 679 US reservoirs

Challenges: construction of informative **features, very large** dataset, …

ResOpsUS:
https://www.nature.com/articles/s41597-022-01134-7#:~:text=ResOpsUS%20is%20the%20first%20multi,reservoirs%20in%20the%20GRanD%20database.

Supervised: Alberto, Matteo Giuliani
Expected time for graduation 9-12 months
Start: in a few months
Contact: **albertomaria.metelli@polimi.it**
matteo.giuliani@polimi.it

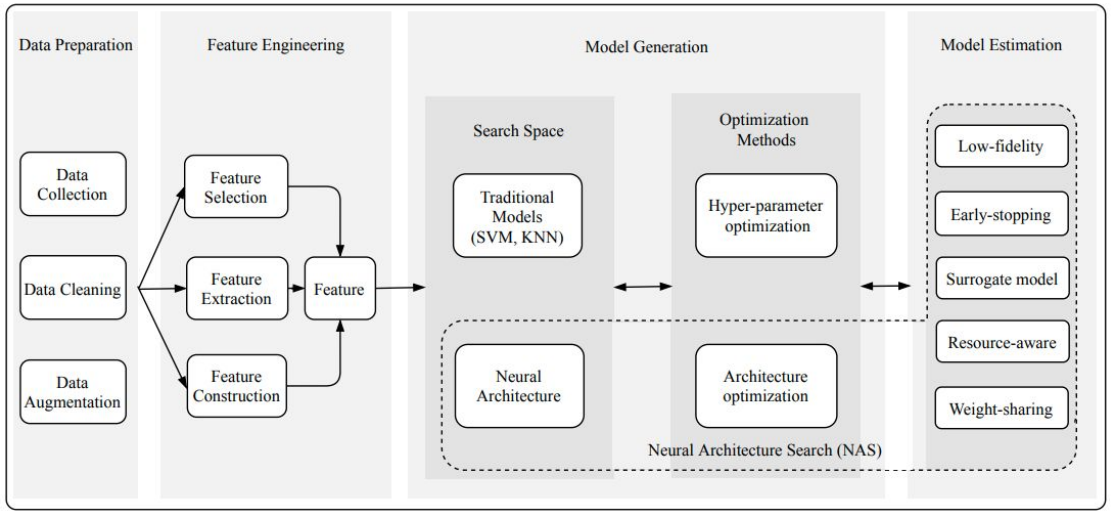
AutoRL

AutoML: Automate all the phases of ML

What if we want to automate the phases of Reinforcement Learning? **AutoRL!!!**

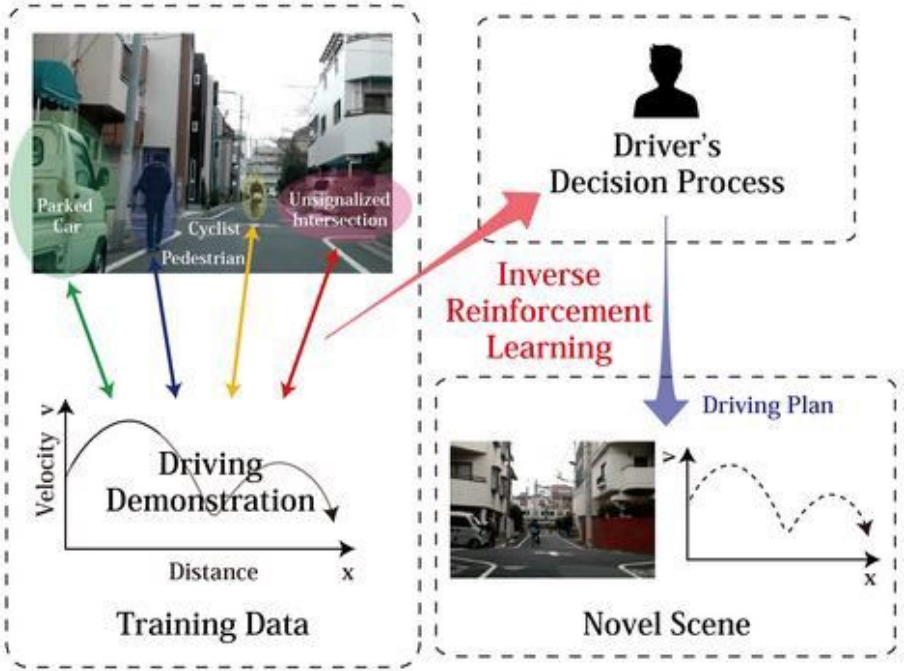
Initial ideas:

- State and action spaces selection
- Reward design
- Algorithm selection
- Evolving the policy space



Supervised: Marco, Alberto, Francesco
Expected time for graduation 9-12 months
Start: now
Contact: `marco.mussi@polimi.it, albertomaria.metelli@polimi.it, francesco1.trovo@polimi.it`

Theoretical Study of Inverse Reinforcement Learning



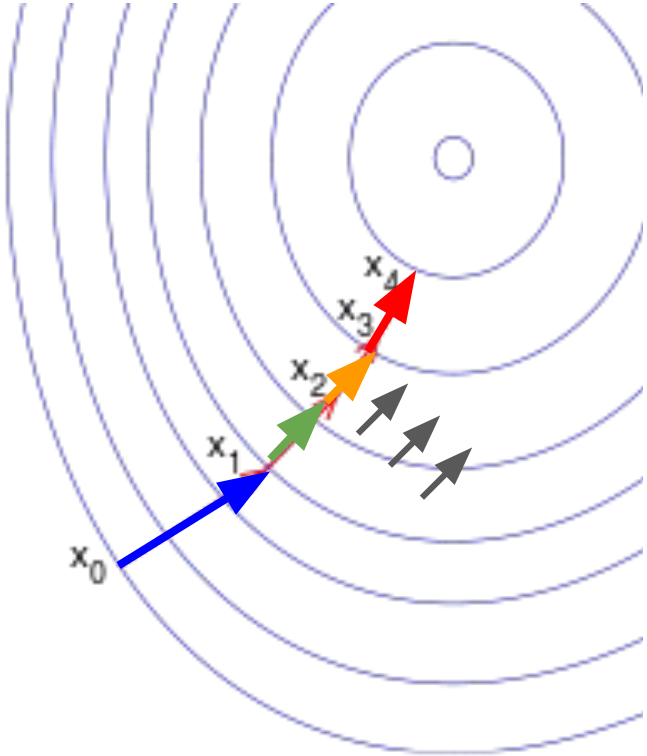
- **IRL**: learn the *reward function* explaining the behavior of an *expert* agent
- **Goal**: study the theoretical properties of IRL
  - What is the minimum number of samples to approximately solve the IRL problem? → **Sample complexity lower bound**
  - Design an **algorithm** matching the sample complexity lower bound

This is an only theoretical thesis!

https://proceedings.mlr.press/v139/komanduru21a.html
https://proceedings.mlr.press/v139/metelli21a.html

Supervised: Alberto e Gianluca
Expected time for graduation 9-12 months
Start: now
Contact: **albertomaria.metelli@polimi.it**

Online Selection of Learning Algorithms



Setting: several learning algorithms optimizing the **same** ML model are available:

- Which one is the most appropriate one?
- Can we find a sequence of methods to get to the optimum point with the smallest number of iterations?

Goal: **select online** to which algorithm assign current data in order to **speed up** finding the optimum

Supervised: Alberto and Francesco
Expected time for graduation 9-12 months
Start: in a few months
Contact: `albertomaria.metelli@polimi.it`, `francesco1.trovo@polimi.it`

I³ LUNG

Personalized medicine for Lung Cancer treatment is the next frontier for medical sciences:

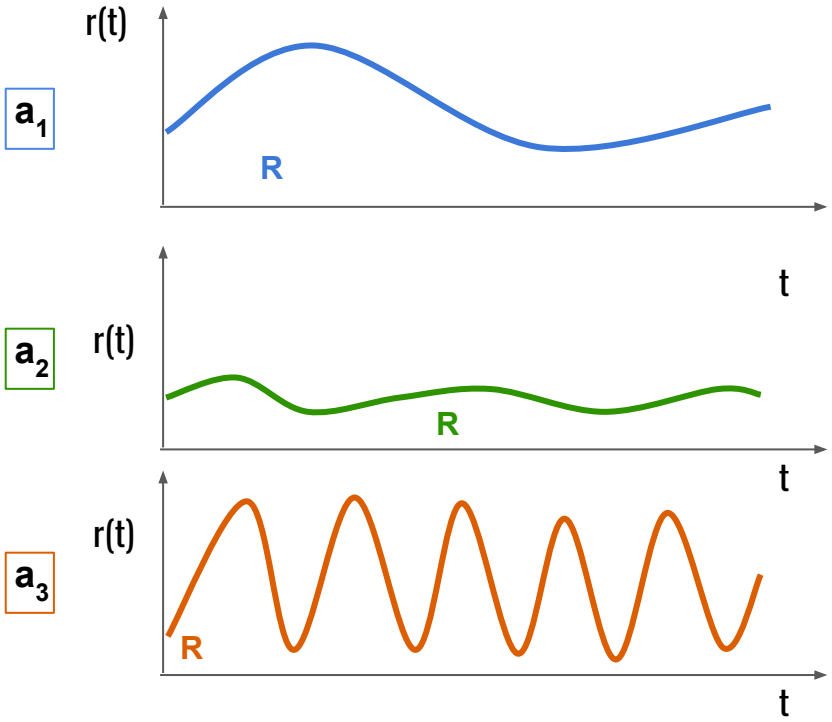
- Allow to select the appropriate treatment for each patient
- Allow to support the physician in his/her choice
- Avoids unnecessary and harmful treatments when ineffective

Applying ML techniques to medical dataset presents several challenges:

- Scarcity of data
- Requirement of interpretation of the result

Supervised by: Francesco
Timespan: 9-12 Months
Start: today
Contact: `francesco1.trovo@polimi.it`

Non-Stationary MAB



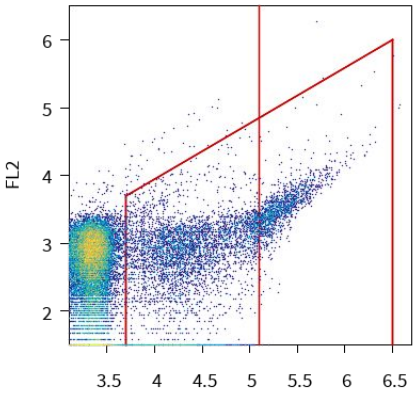
The literature provides already a wide spectrum of algorithms:

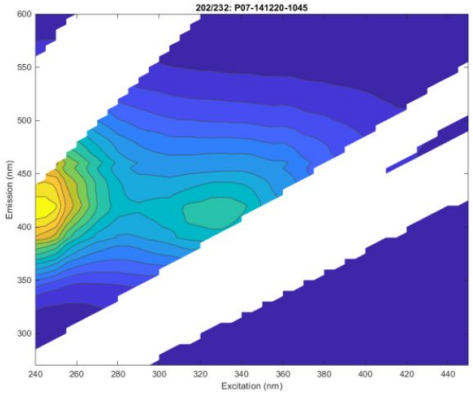
- Active (abrupt changes)
- Passive (smooth changes)

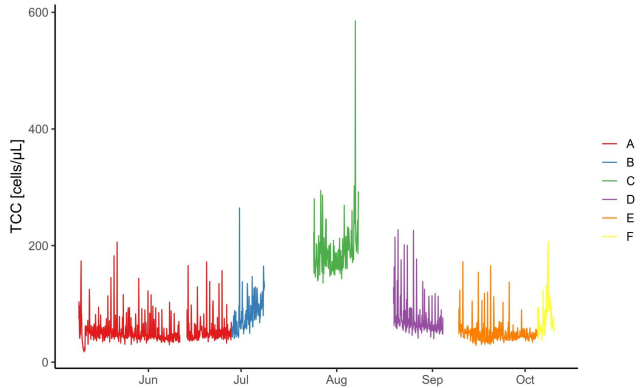
The thesis want to use change point detection techniques to improve the existing algorithms and design smart policies to avoid discarding informative data

Supervised: Francesco e Giacomo
Expected time for graduation 9-12 months
Start: Now
Contact: `francesco1.trovo@polimi.it`

Smart Environmental Monitoring







Determining the best time one should sample a water stream to monitor:

- Contaminants concentrations
- Natural Organic Matter
- Bacterial concentrations

is hard in general due to the non-stationarity of the environment, the complexity of the data and the cost of analysing the samples

Goal: determine the best strategy to identify the time and place where the values of the monitored quantity are the largest/smallest and possibly identifying changes

Supervised: Francesco e Marco G.
Expected time for graduation 9-12 months
Start: Now
Contact: **francesco1.trovo@polimi.it**

Hierarchical RL for Multi-timescale Time Series

Financial time series present a multi-timescale structure, thus, they are non-stationary at multiple levels.

Question: Can we exploit hierarchical RL to effectively deal with multiple timescales?

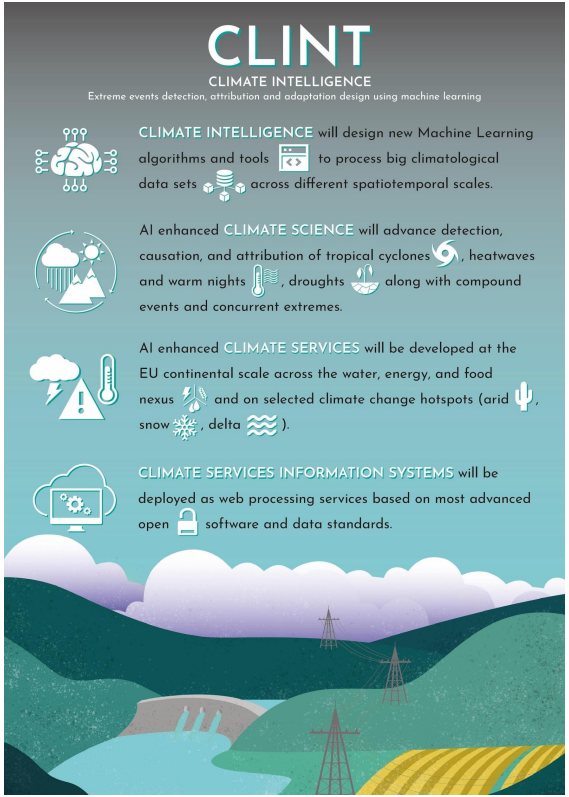
Possible advantages:

- Efficient reuse of samples
- Handling non-stationary behaviours in a decoupled way

Supervised by: Lorenzo (with Luca, Pierre and Edoardo)
Timespan: 9-12 months
Start: immediately
Contact: lorenzo.bisi@polimi.it

CLINT
CLIMATE INTELLIGENCE

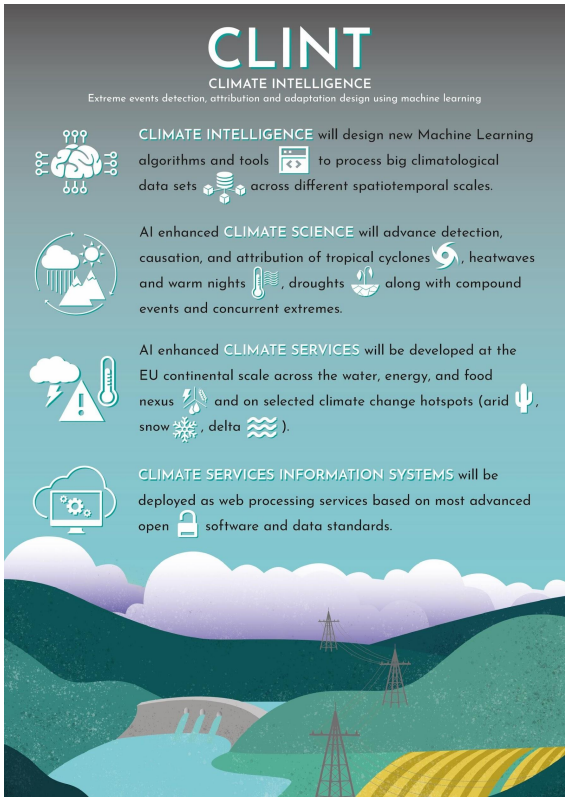
CLINT: Detection



Goal: extreme event (tropical cyclones, droughts) detection using **machine learning**
Data: temporal and spatial high-resolution climate data
Techniques: feature selection, deep learning
Context: 4-year H2020 research project involving 13 partners from France, Germany, Netherlands, Spain, Sweden, UK

Supervised: Paolo, Alberto
Expected time for graduation 9-12 months
Start: now (tropical cyclones), in a few months (droughts)
Contact: **paolo.bonetti@polimi.it**
albertomaria.metelli@polimi.it

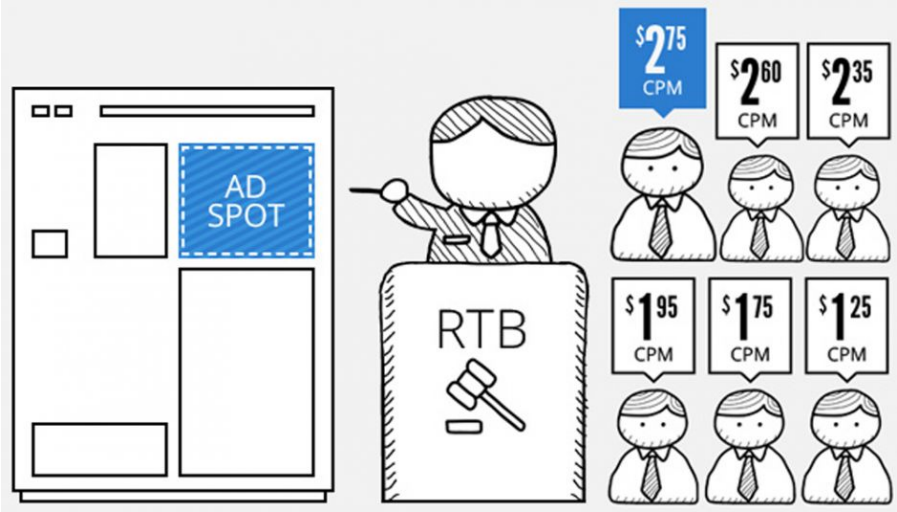
CLINT: Causation



Goal: extreme event causation analysis using **machine learning**
Data: temporal and spatial high-resolution climate data
Techniques: causal inference, causal discovery, deep learning
Context: 4-year H2020 research project involving 13 partners from France, Germany, Netherlands, Spain, Sweden, UK

Supervised: Paolo, Alberto
Expected time for graduation 9-12 months
Start: in a few months
Contact: **paolo.bonetti@polimi.it**
albertomaria.metelli@polimi.it

Real-Time Bidding



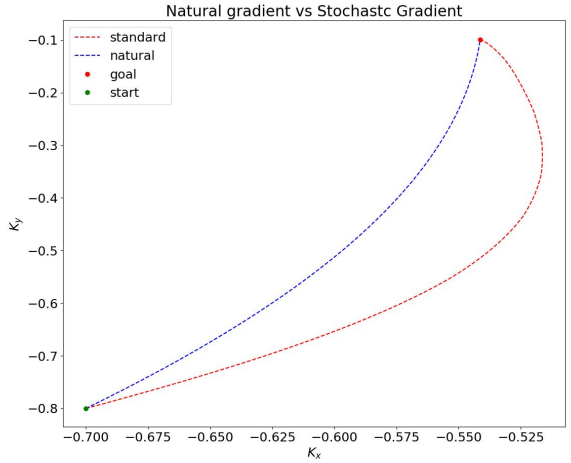
- Every time an online advertising slot is available, an auction is performed to assign it
- The auction is performed in real-time and the advertiser which offers more get the impression
- The objective of this thesis is to design and implement an efficient Real-Time bidding algorithm based on Reinforcement Learning techniques

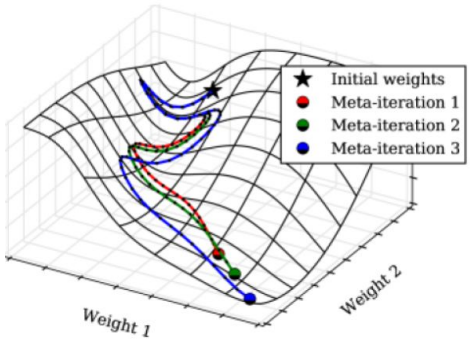
Supervised: Marco
Expected time for graduation: 8-12 months
Start: Immediately
Contact: `marco.mussi@polimi.it`

Meta Reinforcement Learning

Meta Learning: learning to learn better (and faster)

Field of application: hyperparameter and gradient tuning





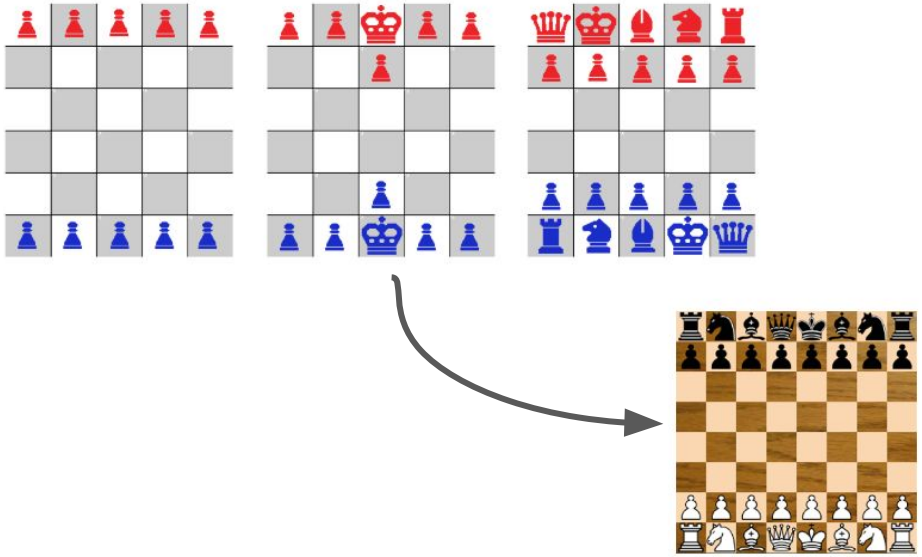
Problems

- How to learn the best hyperparameters?
- Which is the best direction to follow?

Supervised: Luca, Alberto
Expected time for graduation 9-12 months
Start: Now
Contact: `luca.sabbioni@polimi.it`
`albertomaria.metelli@polimi.it`

Curriculum Reinforcement Learning

Curriculum Learning: select a sequence of tasks to improve learning in complex domains



Problems

- How to choose the sequence?
- When should we change task?
- What knowledge should be transferred?
- Theoretical results?

Supervised: Luca, Alberto
Expected time for graduation 9-12 months
Start: Now
Contact: `luca.sabbioni@polimi.it`
`albertomaria.metelli@polimi.it`

Dynamic Control Frequency Adaptation

RL deals with **discrete-time** problems, but the world is **continuous-time** => **time discretization**

Control Frequency Trade-off

High frequency vs *Low frequency*

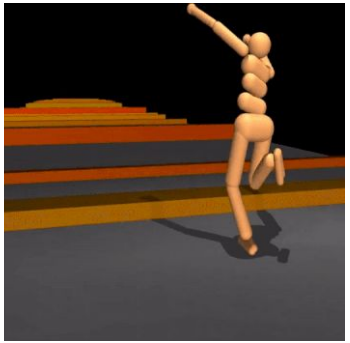
+ more effective control
- overhead
- decreased effect of actions

+ smaller overhead
- less effective control

Previous work: ***Value based*** optimal control frequency

Goal: Gradient Update to learn optimal frequency

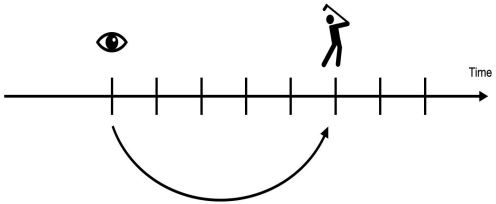
Applications: robot control, trading, ...



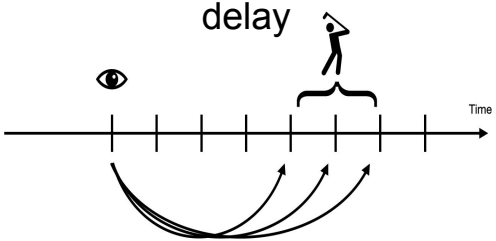
Supervised: Luca, Lorenzo, and Alberto
Expected time for graduation 9-12 months
Start: Now
Contact: luca.sabbioni@polimi.it
lorenzo.bisi@polimi.it
albertomaria.metelli@polimi.it

Delayed RL

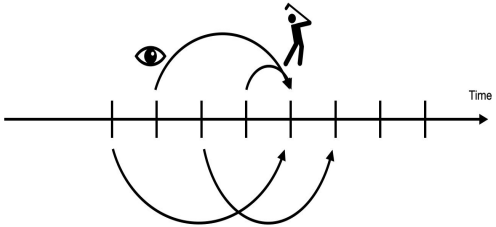
Action delay



Multiple Action delay



State dependent delay



Types of delay:

1. Multiple action execution delays
2. State-dependent delays
3. Delayed reward

Issues:

- Breaks Markov assumption
- Credit assignment problem
- Increased complexity

Goal:

- Define framework (for 1., 2. or 3.)
- Derive theoretical guarantees
- Find practical solution

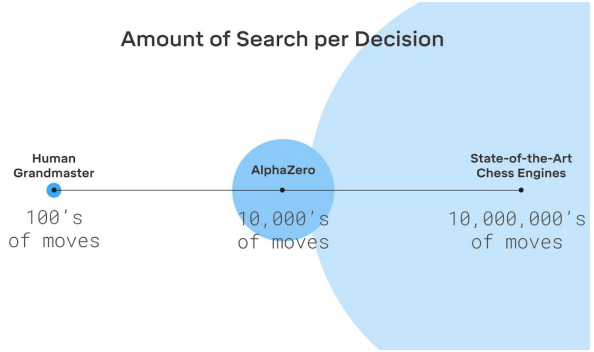
Supervised: Pierre
Expected time for graduation 9-12 months
Start: now
Contact: `pierre.liotet@polimi.it`

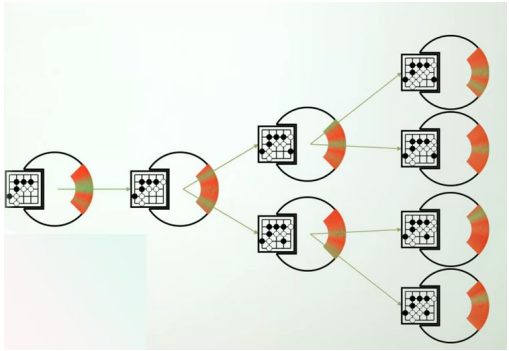
MCTS Under Uncertainty

Reinforcement Learning: learning a task through direct interaction

Planning: Predict environment response and act accordingly

Problem: Combining the two in highly stochastic environments





Application

- Stochastic Games
- Financial Markets
- Navigation Tasks

Supervised: Amarildo
Expected time for graduation 9-12 months
Start: now
Contact: **amarildo.likmeta@polimi.it**

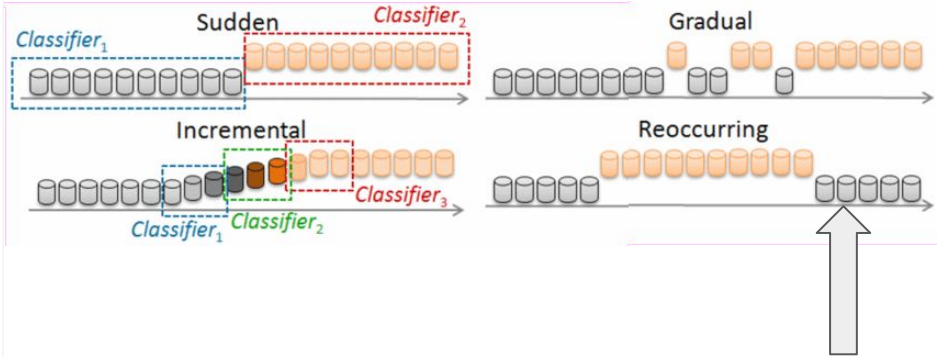
Handling Recurrent Concepts in Evolving Streams

Setting
Streaming data are highly subjective to change over time. However, there may be some regularities in the change and some concepts may reoccur

Issues
Models trained on evolving streams of data need to be constantly updated

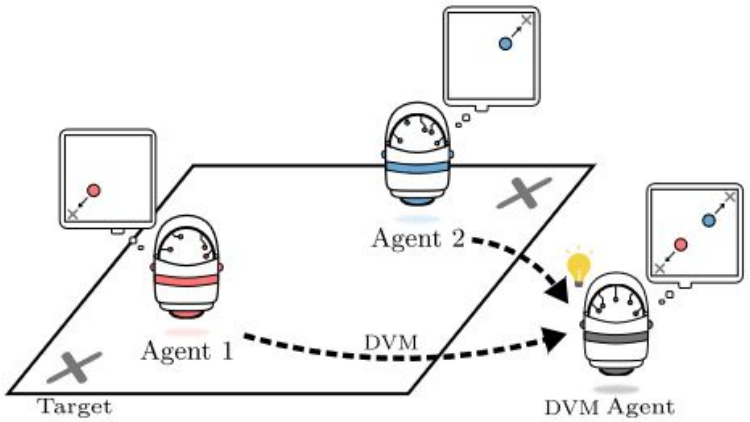
Goals

- Finding sequential patterns of reoccurence between data concepts
- Proactively adapting the model to cope faster with data drifts



Supervised: Alessio
Expected time for graduation 9-12 months
Start: Now / in a few months
Contact: `alessio.russo@polimi.it`

Multi-agent Reinforcement Learning



Problem:
MARL in cooperative and partially observable environments. Multiple agents learning concurrently to cooperate with possibly partial observability (PO) and decentralized decisions

Issues: local non-stationarity, hard effective coordination even without PO, hard scalability to large numbers of agents

Goal: Find new ways to extract useful information and/or guarantee a more sound/scalable learning process

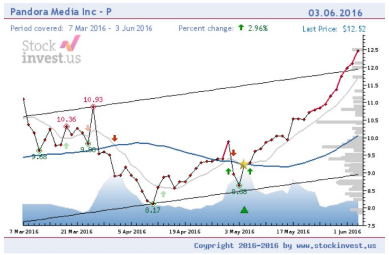
Supervised: Riccardo Z.
Timespan: 9-12 months
Start: in a few months
Contact: `riccardo.zamboni@polimi.it`

Reinforcement Learning in Input-Driven Environments

Setting
In many real-world crucial problems the non-deterministic behavior of the environment depends exclusively on external quantities that are not under the control of the agent.

Examples
- Finance (stock prices)
- Inventory Control (demand)
- Water Reservoir Control (weather)

Goal
How to effectively exploit this peculiar problem structure?



Supervised: Riccardo P.
Expected time for graduation 9-12 months
Start: Now / In a few months
Contact: **riccardo.poiani@polimi.it**

Simulator Selection

Setting

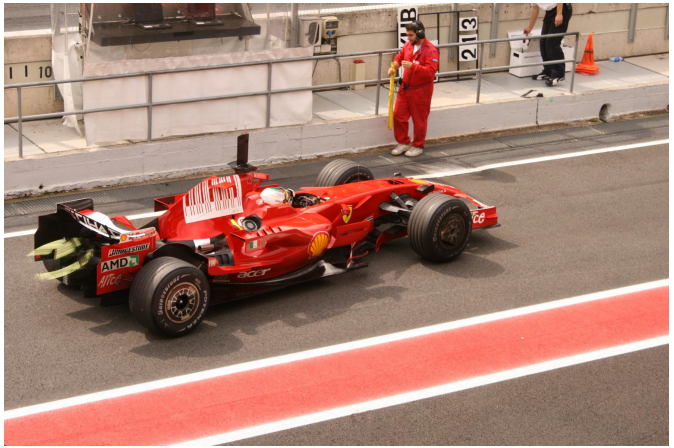
- Two correlated MAB problems
- Pulling arms of the first bandit costs a lot
- Pulling arms in the second bandit costs less
- We are interested in selecting the best arm in the first bandit

Applications

- Select best F1 car setting using different simulators
- Sim2Real Transfer
- Many other applications in which multiple simulators are available

Goal

- How to trade-off costs?
- How to exploit the auxiliary bandit?



Supervised: Riccardo P.
Expected time for graduation 9-12 months
Start: In a few months
Contact: `riccardo.poiani@polimi.it`

Other information

- The usual duration of a thesis ranges from 8 to 14 months (according to the number of credits left and your effort)
- If you are interested in one of these topics, you have to fill this form:
  - https://forms.gle/wc78TbaHUWLV6CR76
- You can find this address on these webpages:
  - http://home.deib.polimi.it/restelli/
  - https://trovo.faculty.polimi.it/
- Deadline: February 28