Theses Proposals

Proff. Alberto Maria Metelli, Marco Mussi

Reinforcement Learning with Provable Stability Guarantees

Advisors: Alberto Maria Metelli, Marco Mussi, Corso Federico

RL

Advantages:

- **Minimal** prior **knowledge** of the **system**.
- Allow to control extremely **complex systems.**

Disadvantages:

- **Not reliable** $\Longrightarrow$ Not for Safety-Critical applications.
- Theory providing safety/stability certificates?

Control Theory

Advantages:

- **Reliable, Rich** and **mature** theory to develop **formal guarantees**.

Disadvantages:

- Knowledge of a model required: complexity limit, **deterministic,** often **linear**.
- **Suboptimal/Conservative** when dealing with **stochasticity, disturbances, unmodelled dynamics.**

Motivating Problem/Research Question?

Can we achieve the “best of both worlds?”

Goal of the Thesis

- Select the most suitable notion of stability in the context of the RL formalism.
- Design Reinforcement Learning Algorithms with provable stability guarantees.

Cost-Aware Reinforcement Learning from Human Feedback

Advisors: Alberto Maria Metelli, Marco Mussi, Diego Alovisetti

Core principle: Human-generated feedback helps RLHF algorithms distinguish *good* actions from *bad* ones (e.g., in the alignment of LLMs and robotics).

↓

Motivating problem/research question:
Human feedback is costly (time, cognitive load, etc.), but training can only use limited resources. *How to optimize*?

Goal of the thesis:
Build a theoretical framework where…

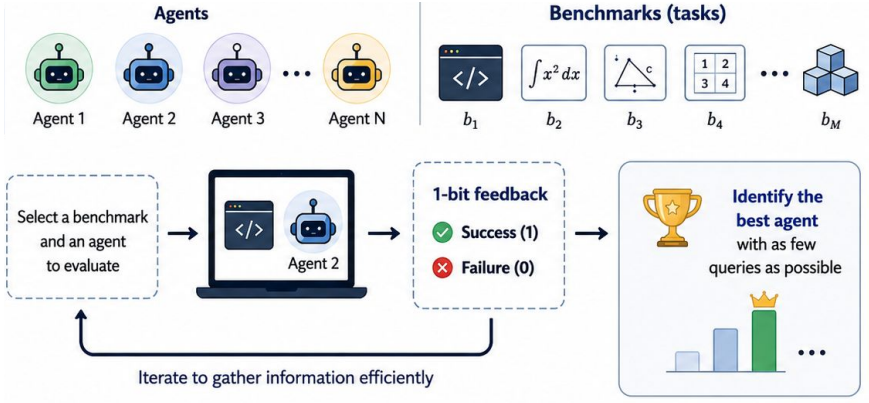
1) A *cost-aware* acquisition function selects feedback signals based on (i) the impact on learning and (ii) the cost of the feedback.

2) We solve the Cost-Aware RLHF problem, achieving successful learning while minimizing the feedback cost.

Active Ranking from 1-bit Feedback

Supervisors: Marco Mussi, Alberto Maria Metelli, Gianmarco Genalti, Cristiano Migali

Motivating Problem. A*rtificial agents* (e.g., LLMs) are increasingly required to solve *complex tasks*. ***Identifying*** the ***best agent*** for a **complex task** (*coding*, *math*) is difficult: humans are *not able* to assign a *numerical score*, they can just tell if the agent *succeeded* or *failed* in a trial (***1-bit feedback***) on given ***benchmarks***. ***Choosing*** the *right benchmark* becomes crucial: it must *neither* be *too easy*, nor *too difficult*.



Research Question. If provided with a *set of benchmarks*, how can we **identify** the best agent for the task through 1-bit feedback, **minimizing** the number of **queries**?

Goal. Given a *mathematical formulation* of the problem:

- *Design* an evaluation ***algorithm*** and provide *theoretical guarantees* on its performance.
- *Analyze* the ***statistical complexity*** of the problem under study.

What you can expect from the thesis

- These are all “Tesi con controrelatore” (max 7 points)
- Approximate duration **9-12 months**
- Working on a cutting-edge research topic with the goal of a **publication in top-tier venues**

What we expect from you

- Strong background in calculus, probability, statistics, linear algebra
- Good programming skills (Python)
- You like design and analyze algorithms

Interested?

If you are interested in one or more of these thesis, please fill this form:

https://forms.gle/DvFmBqeGGHqPDrg76

Deadline: June 9th

Questions? Contact us:
albertomaria.metelli@polimi.it

Slides and links on my personal website:
https://albertometelli.github.io/