Politecnico di Torino
1859

ITALIAN NATIONAL PHD IN ARTIFICIAL INTELLIGENCE

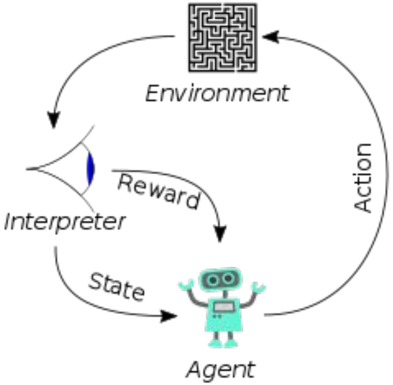
Consiglio Nazionale delle Ricerche

Safe Reinforcement Learning

Mario Fiorino

Reinforcement Learning

Reinforcement learning (RL) is a branch of machine learning where an AI system learns to solve sequential decision-making problems through trial-and-error interactions with its environment.



Agent: agent is a software entity trained to achieve a specific goal

Environment: the world in which the agent operates, performs actions, and gathers feedback signals

Action: a movement that allows to switch from one state to another in the environment.

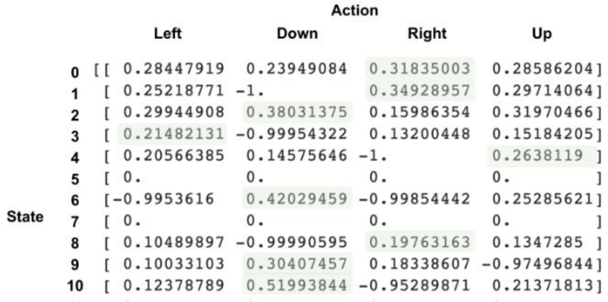
Reward signal: numerical feedback from environment.

Purpose of Reinforcement Learning

Solving a RL task means, after a more or less long exploration of the environment, obtaining a mapping from states to actions (called **policy)**, that maximize the total reward signal over the time.

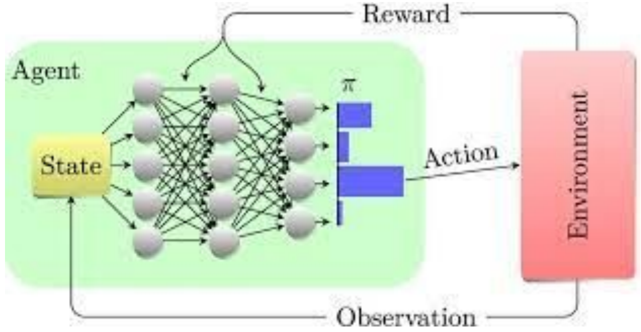
Usually the policy is implemented as:

- Lookup table
- Neural network, whose input is a representation of the state, whose output the probability distribution on possible actions in that state (Deep-RL).



```
                                 Action
                  Left        Down        Right         Up

          0  [[ 0.28447919  0.23949084  0.31835003  0.28586204]
          1   [ 0.25218771 -1.          0.34928957  0.29714064]
          2   [ 0.29944908  0.38031375  0.15986354  0.31970466]
          3   [ 0.21482131 -0.99954322  0.13200448  0.15184205]
          4   [ 0.20566385  0.14575646 -1.          0.2638119 ]
          5   [ 0.          0.          0.          0.        ]
          6   [-0.9953616   0.42029459 -0.99854442  0.25285621]
State     7   [ 0.          0.          0.          0.        ]
          8   [ 0.10489897 -0.99990595  0.19763163  0.1347285 ]
          9   [ 0.10033103  0.30407457  0.18338607 -0.97496844]
          10  [ 0.12378789  0.51993844 -0.95289871  0.21371813]
```



Safe Reinforcement Learning

› What does Safe Reinforcement Learning focus on?

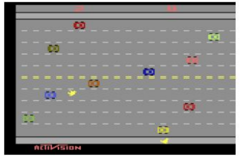
In traditional RL paradigm, typically assumes that agents are free to explore any behavior during learning, with the primary goal of identifying a policy that maximizes long-term rewards. **Safe Reinforcement Learning** (Safe RL) focuses on designing algorithms that optimize performance by maximizing rewards while ensuring adherence to safety constraints.

Roughly speaking, Safe RL is a subset of RL algorithms that, in some way, during training and deployment of the (near) optimal policy, sample “only” safe actions.

Safe Reinforcement Learning

› Why ?

1. Exploit the advantages of the RL methods : ability to learn high-performance control strategies from directly experience (without requiring any prior model).
2. In complex real-world scenarios, creating an accurate simulation to efficiently train an RL agent is often impractical. As a result, training RL agents directly in real-world environments becomes necessary, and **in real-world, safety constraints are a critical issue**



Safe Reinforcement Learning

› Variety of application domains :

1. **Healthcare :** Safe RL can be applied to optimize treatment plans and drug dosages in healthcare. It can help personalize medical treatments by considering individual patient characteristics and adapting to changing patient conditions.
2. **Autonomous Systems and Robotics :** Safe RL can enhance the safety and reliability of autonomous systems and robots; reducing the potential risks involved in these physical systems interacting with the real world.
3. **Power System Control :** Safe RL can train systems deployed in power management to avoid damaging infrastructure.
4. **Finance :** Safe RL algorithms can be employed to develop personalized investment strategies (optimize returns while mitigating the risk of loss)
5. **Human-Machine Interaction :** Safe RL train agents to avoid taking actions that could harm or offend sensibilities of an user (E.g. Recommender systems should not expose users to psychologically harmful).

Safe Reinforcement Learning

› Main challenges in this research :

- **Deal with limitation inherited from the RL :**
  - Sample efficiency : RL algorithms often require a significant number of interactions with the environment to learn optimal policies. This high sample complexity can be impractical or expensive (Curse of Dimensionality).
  - Design of reward functions : Many tasks involve reward that are poorly-defined, or sparse, delayed, or even deceptive! In this case, reward shaping becomes problematic.
  - Generalization : RL agents are trained and evaluated on specific scenarios. This means they often struggle to generalize learned policies to new environments or tasks.

Safe Reinforcement Learning

› Main challenges in this research :

- **Philosophical problems :** How we define "safe" and "unsafe" outcomes. How do we formulate safety specifications to incorporate them into RL? Is this just a task-dependent issue?
- **Conflicting objectives:** Safety constraints and reward maximization may often be conflicting. Actions that maximize reward may not always be the safest.
- **Handling uncertainty:** Algorithms cannot always predict all of the possible consequences of the agent's actions. Visiting a harmful state cannot always be considered excluded a priori. In essence, it is necessary to deal with constraints that we don't know exactly ahead. Under these conditions it can be difficult to ensure safety

Safe Reinforcement Learning

› Current primary approaches (are not mutually exclusive) :

- **Safe optimization algorithms** : This methods typically involves maximizing the return keeping other types of measures (such as cost function that measures the degree of safety given a specific state-action pair) within some specified bounds. E.g., Constrained RL.

- **Safe exploration - external knowledge** : Incorporate into the learning process some external knowledge that guides the agent to prevent choosing unsafe actions. Typically this could be an initial information specifying unsafe regions to avoid; or, it could be a human or other decision-making entity (e.g., an ANN) advising the exploring agent in real time.

- **Safe exploration - external risk metric** : Define a risk measure that is used to determine the probability of selecting an unsafe actions during the exploration process (while the classic RL process substantial remains unchanged). Essentially, it works as a filter for unsafe actions.

Conclusion

Safe RL has the potential to bring many positive impacts in different fields of our society, by introducing adaptive systems that can tailor their actions to complex and uncertain environments while minimizing the risk of damage.

Thanks for the attention

Useful References

- Sutton, Richard S. and Barto, Andrew G., Reinforcement Learning: An Introduction, MIT Press
- Shangding Gu, Long Yang, Yali Du, Chen, Walter, Wang, Yang, Knoll, A Review of Safe Reinforcement Learning: Methods, Theory and Applications, arXiv:2205.10330
- Garcıa, Javier, and Fernando Fernández. A comprehensive survey on safe reinforcement learning. Journal of Machine Learning Research 16.1
- Weiye Zhao, Tairan He, Rui Chen, Tianhao Wei, Changliu Liu. State-wise Safe Reinforcement Learning: A Survey. arXiv:2302.03122
- Youngmin Kim, Richard Allmendinger, Manuel López-Ibáñez. Safe Learning and Optimization Techniques: Towards a Survey of the State of the Art. arXiv:2101.09505
- Paul Christiano, Jan Leike, Tom B. Brown, Shane Legg, Dario Amodei; Deep reinforcement learning from human preferences. arXiv:1706.03741

License