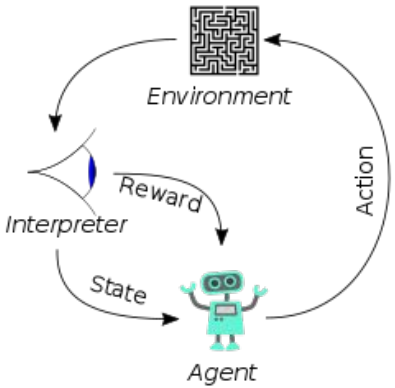
Safe Reinforcement Learning

Mario Fiorino

NATIONAL AI PhD FALL SCHOOL, Torino, November 20-22, 2023

Reinforcement Learning

RL is a type of machine learning in which an agent (or AI-driven system) learns to solve sequential decision problems, through trial-and-error interactions with an environment.



Agent: a software to train for a certain purpose.

Environment: the world in which the agent can stay and performs actions and a collection of state signal.

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

Reward signal: numerical feedback from environment.

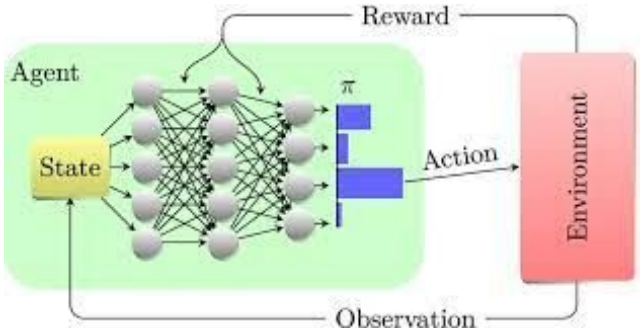
Purpose of Reinforcement Learning

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 (e.g. of Q-value),
- Neural network, whose input is a representation of the state, whose output the probability distribution on possible actions in that state.(DeepRL field)

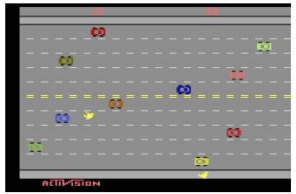
		Action			
		Left	Down	Right	Up
State	0	[[0.28447919	0.23949084	0.31835003	0.28586204]
	1	[0.25218771	-1.	0.34928957	0.29714064]
	2	[0.29944908	0.38031375	0.15986354	0.31970466]
	3	[0.21482131	-0.99954322	0.13200448	0.15184205]
	4	[0.20566385	0.14575646	-1.	0.2638119]
	5	[0.	0.	0.	0.]
	6	[-0.9953616	0.42029459	-0.99854442	0.25285621]
	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 it mean?

Safe Reinforcement Learning (Safe RL): Reinforcement Learning methodologies that balance the maximization of a reward (i.e. performance optimization) with the safety constraints for the agent and the surrounding environment



› Why ?

1. Exploit the advantages of the RL methods : ability to learn from directly experience (without requiring any prior model) in complex and various scenario.
2. If we apply RL techniques in a simulated environment, our only problem is related to computational resources; but, **in real-world applications safety constraints are a critical issue**

Safe Reinforcement Learning

› Variety of application domains :

1. **Healthcare and Well-being** : Safe RL can help personalize medical treatments by considering individual patient characteristics and the evolution of the disease.
2. **Autonomous Systems and Robotics** : Safe RL algorithms empower robots to operate autonomously in uncertain and dynamic environments, mitigating safety risks. In more, this approach reduce dependency on manual planning.
3. **Finance** : Safe RL can help to modeling and mitigating risks in dynamic markets, contributing to financial stability.
4. **Resource Management** : Safe RL can minimize waste and conserve critical resources.
5. **Human-Machine Interaction** : By incorporating safety constraints, SafeRL agents can adapt their interaction to the user's preferences, and avoid actions that could harm or send away of an user (Use case: recommender systems).

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 goals that are complex, poorly-defined, or hard to specify.
  - Continuous learning and adaptation : RL algorithms should be able to integrate acquired knowledge, and continuously learn and adapt to changing environments.
  - Limited theoretical understanding : There is still limited theoretical understanding of RL, which can make it difficult to design new algorithms and improve the performance of existing ones.

Safe Reinforcement Learning

› Main challenges in this research :

- **Formalization of the safety concept:** Defining what constitutes safety can be challenging, especially in complex and dynamic environments (Note: different safety constraints may apply depending on the specific task and the desired level of risk tolerance).
- **Balancing safety and reward :** Safe RL algorithms must learn to balance the goals of maximizing reward and ensuring safety (E.g. Safe RL systems designed to prioritize safe actions may limit its ability to explore and discover optimal strategies). This can be difficult, as the two goals may often be conflicting.
- **Handling uncertainty:** Safe RL algorithms cannot always predict all of the possible consequences of the agent's actions, the visiting a risky state cannot always be considered excluded a priori. In essence, it is necessary to deal with constraints that we don't know exactly ahead.

Safe Reinforcement Learning

› Current main approaches :

- **Safe optimization algorithms :** make safety aspects part of the policy itself. Modify optimization criteria (i.e. the performance measure to guide the agent's learning) to include constraints that ensure the policy parameters stay within a safe region.
- **Safe exploration - incorporate external knowledge :** modify the exploratory process to prevent the agent from choosing unsafe actions during training. This can be done by providing an initial information (or a model) specifying unsafe regions to avoid. Alternatively, it could be a policy derived from a finite set of demonstrations provided by a human or another decision-making entity (e.g. ANN).
- **Safe exploration - risk driven :** in contrast to approach to optimization above, this risk metric is not incorporate constraints during policy formulation, but it serves to guide the agent in the exploration phase, to avoid risky situations while learning.

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