Reinforcement Learning

Reinforcement learning (RL) has become an increasingly popular and powerful tool in machine learning due to its ability to learn from experience and adapt to changing environments. It can be seen as another area of machine learning that has several differences when compared to supervised or unsupervised learning. In RL, the agent is not provided with labeled data (different to supervised learning); however, using rewards it is told whether progress is made (different to unsupervised learning). By using an optimization technique such as policy gradient or a genetic approach, and finding a "balance between exploring the environment, looking for new ways of getting rewards, and exploiting sources of rewards that it already knows", the agent's goal is to find a good policy, i.e. an algorithm that maximizes rewards over time (Géron, 2020). A further important difference between (un)supervised learning and RL is that consecutive observations are not independent. The algorithm remains in regions of the environment; thus, correlating observations. The RL algorithm selects an "action" by summing all rewards this action has led to in the past and possibly introducing a discount factor for more distant rewards.

Applications for RL can be found in a variety of fields, including as robots, gaming, banking, and healthcare. Its capacity to learn the best policies in complicated contexts, or applications, is one of its key benefits. For instance, RL can be used in robotics to teach a robot how to control a task, such gripping an object, moving, or flying a drone. RL can also be used in online personalization, such as choosing what video or music to play next. The algorithm receives a reward in the form of the user listening to the song / watching the video the algorithm selected (Géron, 2020).

RL's ability to adjust to changing environments is another advantage. RL agents are highly suited for dynamic contexts because they can learn from their experiences and modify their behavior accordingly. For instance, in the financial industry, where market circumstances are continuously shifting, RL might be used to trade stocks or manage investment portfolios. As another example, RL can also be used in the healthcare industry to tailor treatment programs for patients depending on their particular health traits and evolving conditions. In this domain, it can for example be introduced for automated medical diagnosis or drug discovery, design, and development (CapeStart, 2021).

However, the applications of RL also include some challenges, not least to mention, the credit assignment problem. If there is a large delay between an action the algorithm takes and the resulting reward, the agent has problems assigning the reward to the respective action. Thus, a main challenge of RL is the careful design of the reward function. If the reward function is poorly designed, the agent may learn suboptimal policies or even harmful behaviors. Careful consideration must be given to the design of the reward function to ensure that the agent learns the desired behavior.

Another challenge worth mentioning is RL's need for large amounts of data and computation. It requires a significant amount of data to learn a good policy, which can be costly or time-consuming to collect. As the training in the real-life setting can be hard and slow, generally a simulated environment is needed for training the RL algorithm. This, however, also introduces the risk of moving the agent from the simulated environment to the real-life setting and accompanying unexpected/not covered changes the agent is faced with. Further, RL algorithms require significant computational resources, making it challenging to scale to large problems.

Finally, while one mentioned beneficial application of RL is healthcare, it can also be challenging in such safety-critical applications, where mistakes can have severe consequences. Only if the RL algorithm has been properly tested and the safety of it carefully evaluated, it can be deployed in practice; thus preventing accidents or harm to individuals.

In conclusion, RL is a powerful tool that has found applications in a wide range of domains. In complex and dynamic contexts, RL has the ability to learn the best courses of action, but it also faces significant difficulties, such as the requirement for a lot of data and careful reward function design. It is crucial to carefully weigh the advantages and difficulties of applying RL as it develops in order to make sure that it is done so safely and successfully.

CapeStart (2021) *Reinforcement learning in health care: How it can help*, *CapeStart*. Available at: https://www.capestart.com/resources/blog/reinforcement-learning-in-health-care-why-its-important-and-how-it-can-help/ (Accessed: April 11, 2023).

Géron Aurélien (2020) *Hands-on machine learning with scikit-learn, Keras, and tensorflow concepts, tools, and techniques to build Intelligent Systems*. Beijing: O'Reilly.