Model-based off-line reinforcement learning

Giovanni Montana

Reinforcement learning (RL) is a branch of machine learning concerned with optimising sequential decision making in dynamic environments [1]. RL problems are framed as Markov Decision Processes (MDPs) in which an agent tries to learn the best series of actions (policy) which maximises some environmental reward signal for a particular task, where each action is based on some (temporal) state representation of the environment. Deep reinforcement learning (DRL) extends RL to higher dimensional state and action spaces via the use of neural networks, allowing much more complicated tasks to be undertaken, such as continuous control problems found in robotics/autonomous driving, as well as learning policies purely from pixel representations of states (i.e. images/video) [2].

DRL algorithms are categorised as either model-free or model-based. Model-free algorithms learn policies purely through sampled interactions with the environment, whereas model-based approaches use these samples to also build a model of the environment, learning policies from both real and model-generated samples. As such, model-based algorithms are typically more sample efficient than model-free, however their asymptotic performance (in terms of rewards) is poorer due to errors induced by model approximation.

Whether model-free or model-based, DRL requires online interaction with the environment as part of the data collection and learning process. This limit DRL application in settings where data collection is time-consuming, expensive and/or dangerous. In areas such as robotics and autonomous driving this can be partially alleviated through the use of simulators, but in settings such as healthcare this potentially becomes a major obstacle.

This has led to the field of offline-DRL, which attempts to apply the principles of RL to fixed datasets, with further interaction with the environment prohibited. This generates new challenges however, chief being how to assess actions absent from the data (technically known as distributional shift) [3]. Approaches to date focus on adapting existing model-free and model-based algorithms to the offline setting using ensemble techniques [4], policy constraints [5] and conservative estimation of action quality [6]. These approaches have produced encouraging results on simulated datasets [7], paving the way for real-world applications in complex settings, such as healthcare.

The goal of this project is to introduce the student to offline-DRL, with a view to algorithm development and real-world application as part of a subsequent PhD. The student will compare and contrast existing model-based offline-DRL approaches, implement algorithms in code, and use this knowledge to initiate ideas to take forward to PhD. The PhD student will be part of a larger team that is being formed as part of a 5-years Turing AI Fellowship and clinical data will be provided by NHS partners to support the project.

References

[1] Sutton, R. S. & Barto, A. G., 2018. *Reinforcement learning: An introduction.* 2nd ed. Cambridge, MA: The MIT Press

[2] François-Lavet, V. et al., 2018. *An Introduction to Deep Reinforcement Learning.* s.l.:Now Foundations and Trends

[3] Levine , S., Kumar , A., Tucker , G. & Fu, J., 2020. Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems. *arXiv*

[4] Agarwal , R., Schuurmans , D. & Norouzi, M., 2019. *An Optimistic Perspective on Offline Reinforcement Learning.* International Conference on Machine Learning

[5] Fujimoto , S., Meger , D. & Precup, D., 2019. *Off-Policy Deep Reinforcement Learning without Exploration.* Proceedings of the 36th International Conference on Machine Learning

[6] Kumar , A., Zhou , A., Tucker , G. & Levine, S., 2020. Conservative Q-Learning for Offline Reinforcement Learning. *arXiv*

[7] Fu , J. et al., 2020. D4RL: Datasets for Deep Data-Driven Reinforcement Learning. *arXiv*