Exploration in Reinforcement Learning

Adam Jedlička

Department of Adaptive Systems, Institute of Information Theory and Automation, Pod Vodárenskou věží 1143, 182 00 Prague 8, Czech Republic

e-mail: jedlicka@utia.cas.cz

Key words: Reinforcement Learning, Exploration methods, Markov Decision Processes

A wide variety of tasks including modeling biological problems can be modeled by Markov Decision Pro- cess (MDP). MDP consists of an agent interacting with the environment. The agent observes the environ- ment state and influences it by purposefully selected actions.

Reinforcement learning (RL) is an approach to solving MDP. In RL the agent learns to make the optimal actions by using feedback (reinforcement) signal. Due to its ability to handle dynamic environments with high uncertainty, RL has been successfully applied to various biological problems: generating novel mo- lecular structures in drug discovery [Liu, K., et al. (2018). “Deep Reinforcement Learning for de Novo Drug Design.”], predicting protein folding [John Jumper et al. (202) “Highly accurate protein structure prediction with AlphaFold”], etc.

A very basic application of MDP terms in the example of the task of discovering new drugs mentioned ear- lier is as follows. A generative model (agent) learns a series of actions to create new molecules (states) for maximizing a score given by a predefined score function. RL is applied similarly in an example of genome assembly and other tasks in biology. The exploration is an important task in RL that i) enables the agent to efficiently discover new “state-action” pairs; ii) improves the agent’s ability to adapt to new, unseen, situations; iii) helps the agent to generalize knowledge acquired.

The so-called exploration-exploitation dilemma refers to optimizing the trade-off between discovering new states (exploration) and using already gathered knowledge for immediate reward The importance of the proper choice of this exploration algorithm lies in the potentially large improvement in the speed of convergence of the RL algorithm. The choice of a well-performing exploration algorithm is task and do- main-specific thus there is no universal algorithm that would perform the best for every given task. The proposed poster will i) briefly introduce a mechanism of how RL works along with the comprehensi- ve implementation of a biology-related task into an MDP that is suitable to be solved by RL. ii) describe several exploration algorithms (from rather simple ε-greedy exploration to more complex methods such as the Intrinsic Curiosity Module (ICM)) along with their benefits and show how exactly they fit into the overall RL mechanis.