Dashboard for the Visual Analysis of Reinforcement Learning Environments

Tiago Araújo
tiagodavi70@ua.pt
João Alves
jbga@ua.pt
Paulo Dias
paulo.dias@ua.pt
Beatriz Sousa Santos
bss.@ua.pt

IEETA - DETI
University of Aveiro
Aveiro, PT

Abstract

Modern reinforcement learning is a field its with many dynamic and fruitful integration with various engineering and scientific disciplines. However, it comes with an inherent challenge—the understanding of its models, which makes it challenging for humans to trust the decisions made by these algorithms. In response to this challenge, we propose the development of an interactive dashboard designed to ease the analysis of reinforcement learning environments. This dashboard offers a comprehensive set of features to visualize critical elements of reinforcement learning experiments, encompassing agent behavior, reward dynamics, and exploration of time-based features.

1 Introduction

Reinforcement learning (RL) is a technique that enables machines to learn and make decisions in complex environments through direct experience with the environment. One of the most exciting aspects of modern reinforcement learning is the substantive and fruitful interactions with other engineering and scientific disciplines [6]. However, one of the inherent challenges of reinforcement learning is the lack of interpretability of its models, making it difficult for humans to understand and trust the decisions made by these algorithms.

A human can help address some of the challenges associated with RL model development, such as the risk of bias or errors in the data or the model, as well as the difficulty of interpreting complex models. A constant human supervision is needed in many applications and using interpretable methods one can easily automatize this analysis [1, 9]. We propose an interactive dashboard[1] designed to ease the analysis of RL environments. It offers a rich set of features for visualizing important aspects of RL experiments, including agent behavior, reward dynamics, and state-space exploration.

In this context, visualization techniques can help to present explanations of models and predictions in a more understandable, explainable and interpretable way, especially for those who are not experienced in the field of RL. Even for those who are experienced, it can help with many human aspects of usage of these methods, like fairness, transparency and accountability. Using aspects of human vision, one can tamper with models and uncover how they work. In recent years, explainable RL has been gaining relevance as a research field offering unique opportunities [2, 10]. As the field evolves, major research firms like DeepMind and OpenAI are already exploring it, applying it to better understand the behaviour of their in-house developed models [3, 7].

2 Reinforcement Learning

Reinforcement learning (RL) is a distinct type of machine learning that emphasizes identifying the optimal course of action in diverse circumstances to attain the maximum possible reward. Unlike other learning techniques, the learner isn't provided with explicit instructions on which actions to take. Instead, they must explore and deduce the actions that lead to more significant rewards. In general, the actions taken may affect both immediate and future rewards. The three main elements of RL are the environment, a policy and a reward signal.

In the field of RL, the environment refers to the external system that an agent interacts with in order to improve its decision-making abilities. It is crucial to understand the environment as it provides the necessary context for the agent to take actions, receive feedback, and learn how to maximize its cumulative rewards over time. Typically, the environment is formalized as a Markov Decision Process (MDP), which consists of state and action spaces. The state space comprises all possible states that the environment can be in, and these states contain all relevant information about the environment's current situation. The action space, on the other hand, is the set of all possible actions that the agent can take. A policy serves as a guide for a learning agent's actions in any given situation. It outlines the appropriate course of action based on the agent's perception of the environment. As the policy governs the behavior of a reinforcement learning agent, it is a critical component of its functioning.

Reinforcement learning relies on a reward signal to guide the agent towards its objective. This signal takes the form of a single number, known as the reward, which the environment sends to the agent at each time step. The agent's ultimate aim is to maximize its cumulative reward over time. This signal serves as the agent's compass, distinguishing positive from negative events. The reward signal is pivotal in refining the policy; if the policy results in a low reward, the agent may opt for an alternate action in analogous situations in the future.

An episode refers to a finite sequence of time steps during which the RL agent engages with the environment. Each episode offers a distinct illustration of how the agent interacts with the environment, starting from an initial state and progressing through a series of state-action-reward transitions until a stopping condition is met. Episodes are particularly valuable in RL settings where there is a definitive sense of accomplishment or closure in the task or environment. By adopting an episodic structure, agents can refine their decision-making approaches over multiple interactions with the environment, leading to enhanced performance in cumulative rewards or task completion.

2.1 Environment

Our work leveraged the Lunar Lander [2] environment to develop the proposed platform. This environment presents a classic rocket trajectory optimization problem where the objective is to land the lander safely on the ground. The observation space comprises an 8-dimensional vector consisting of the lander's coordinates in the x and y axes, its linear velocities in the x and y axes, its angle, its angular velocity, and two booleans indicating if each leg is in contact with the ground. The action space involves four discreet actions: no action, firing the left orientation engine, firing the main engine, and firing the right orientation engine. While the starting position of the lander remains constant, we apply a random initial force to its center of mass.

3 Information Visualization

Visual representations offer users the ability to uncover insights and make informed decisions. They facilitate the creation of presentations that highlight patterns, whether they involve trends, groups, intervals, or outliers, as well as individual data items. Employing computational support to visualize and interact with abstract data enhances human cognition, empowering users to acquire knowledge about the data and its interconnectedness [8].

Unlike static media, computer systems offer a distinct advantage in dynamic visualization. Visualization software continually constructs vi-

[1] `https://github.com/tiagodavi70/lunarlander`

[2] `https://gymnasium.farama.org/environments/box2d/lunar_lander/`

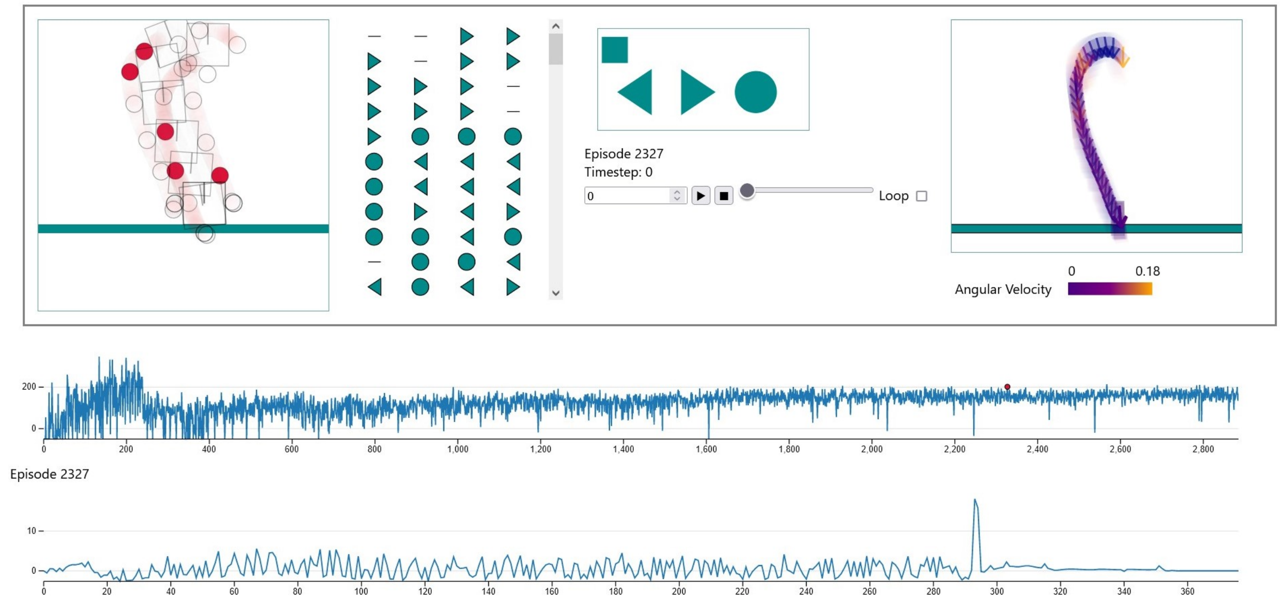

Figure 1: Dashboard for analysis of RL models on the Lunar Lander environment. The dashboard features at the top an agent visualization during an episode with actions marked in red, an list paired with a joystick like view, to highlight an action of the agent in a timestep, a player that manipulates the animation of time based visualization and the angular velocity view of the episode. At the bottom it features total rewards of the model and a single episode rewards view of the agent.

sualizations that evolve as the analysis progresses through user interactions. These interactions, driven by user actions, trigger changes in the visual displays [4]. Researchers [4, 5] have identified several crucial interaction tasks in this context, including configuring visualization designs, defining visual encodings, classifying visual elements, locating specific values, recovering values like maximum and minimum, and uncovering correlations between different data items.

3.1 Dashboard

Our proposed solution for the problem of RL interpretability is a dashboard for the Lunar Lander environment models. This visual idiom allows insights through episode and actions analysis, and it is shown in Figure 1. The dashboard features a set of visualizations covering many important aspects of RL problems: time manipulation, interaction and brushing. Nonetheless, most features presented in this work can easily be transposed to other 2D scenarios. In RL interpretability, it is crucial to have episode and timestep analysis of the model, showing how the experience is gained during training. The dashboard features time analysis with a player, that enables users to see parts of the episode animated in two views: the agent view and the velocity view.

Interaction in visualization refers to the ability of users to engage with visual representations actively. Through user-driven actions such as clicking, hovering, or dragging, interactive visualizations offer dynamic exploration and manipulation of data. This enables the user to build a chart unique for their insights. It also presents for the user how the selection of a episode impacts the visualization of the agent, from a overview perspective of all episodes to a specific one.

Brushing is an interactive technique used in visualization to highlight and select data points or regions. By "brushing" over the visual elements, users can visually emphasize certain parts of the data, which can then trigger updates in linked views or provide context for further analysis. This is imperative for analysis of RL models, as the volume of data generated through training or execution of the model can be massive. The dashboard allows brushing of episode based information and allows manipulation of timesteps.

4 Results, Discussion & and Future Work

The main takeway of this work is the usage of simple visualization methods to interpret RL models results. For example, in this version of the dashboard is already possible to see a pattern of stability: the model learns to be stable with rewards near its objective (200), excluding risky moves that could reward more (> 300) but also fail catastrophically (< 100).

Future works include conducting user studies involving experts from the fields of RL and Visualization Information fields to evaluate the tool usability and effectiveness. We also will add other temporal visualization methods to illustrate decision evolution over time and include interactive techniques to compare different models, or the same model in different training stages.

References

[1] Gulsum Alicioglu and Bo Sun. A survey of visual analytics for explainable artificial intelligence methods. *Computers & Graphics*, 102:502–520, 2022.

[2] Leila Arras, Ahmed Osman, and Wojciech Samek. Clevr-xai: A benchmark dataset for the ground truth evaluation of neural network explanations. *Information Fusion*, 81:14–40, 2022.

[3] Jacob Hilton, Nick Cammarata, Shan Carter, Gabriel Goh, and Chris Olah. Understanding rl vision. *Distill*, 2020. doi: 10.23915/distill.00029. https://distill.pub/2020/understanding-rl-vision.

[4] Tamara Munzner. *Visualization analysis and design*. CRC press, 2014.

[5] Robert Spence. *Information Visualization: An Introduction*. Springer, 2014.

[6] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. A Bradford Book, Cambridge, MA, USA, 2018. ISBN 0262039249.

[7] Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, Nov 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1724-z.

[8] Colin Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2019.

[9] Lindsay Wells and Tomasz Bednarz. Explainable ai and reinforcement learning—a systematic review of current approaches and trends. *Frontiers in artificial intelligence*, 4:550030, 2021.

[10] Xuhai Xu, Anna Yu, Tanya R Jonker, Kashyap Todi, Feiyu Lu, Xun Qian, João Marcelo Evangelista Belo, Tianyi Wang, Michelle Li, Aran Mun, et al. Xair: A framework of explainable ai in augmented reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–30, 2023.