Contribution ID: **52**   Type: **Poster**

Reinventing Astronomical Survey Scheduling with Reinforcement Learning: Unveiling the Potential of Self-Driving Telescopes

The rise of cutting-edge telescopes such as JWST, the Large Synoptic Survey Telescope (LSST), and the Nancy Grace Roman telescope (NGRT) has introduced a new era of complexity in the realm of planning and conducting observational cosmology campaigns.

Astronomical observatories have traditionally relied on manual planning of observations, e.g., human-run and human-evaluated simulations for every observing scenario, which can potentially result in suboptimal observations in terms of scheduling optimization.

Reinforcement learning (RL) has been well-demonstrated as a valuable approach for training autonomous systems, and it may provide the basis for self-driving telescopes capable of scanning the sky and collecting valuable data.

We have developed a framework for statistical learning-based optimization of telescope scheduling that can enhance data acquisition given a predefined scientific reward, e.g., optimizing for a volume-based survey.

The observational campaign is framed as a Markov Decision Process (MDP), a mathematical framework that effectively captures the essence of sequential decision-making.

We compared several RL algorithms applied to a simulated offline dataset, i.e., pre-recorded interactions between the telescope and the sky, considering a discrete set of sky locations the telescope is allowed to visit.

In our study, we conducted comparisons between policy-based, value-based methods, and evolutionary computation strategies.

Value-based methods, and in particular Deep Q-Networks (DQNs), have shown remarkable success in the optimization of astronomical observations for our dataset.

Our experimental results on the test set, demonstrate that the combination of dataset preprocessing techniques, along with the combination of well-known improvements from the literature, such as Dueling DQN, n-steps Bellman unrolling, and noisy networks, yield high performances and capabilities to generalize on unseen data for our task.

In the full environment, the average reward value in each state was found to be 92%±5% of the maximum possible reward. The results from the test set showed an average value of 87%, with a standard deviation of 9%.

Primary author: TERRANOVA, Franco (University of Pisa, Fermilab)

Co-authors: NORD, Brian (Fermilab); NEILSEN, Eric (Fermilab); VOETBERG, M (Fermilab)

Presenter: TERRANOVA, Franco (University of Pisa, Fermilab)

Session Classification: Posters

Track Classification: Paris