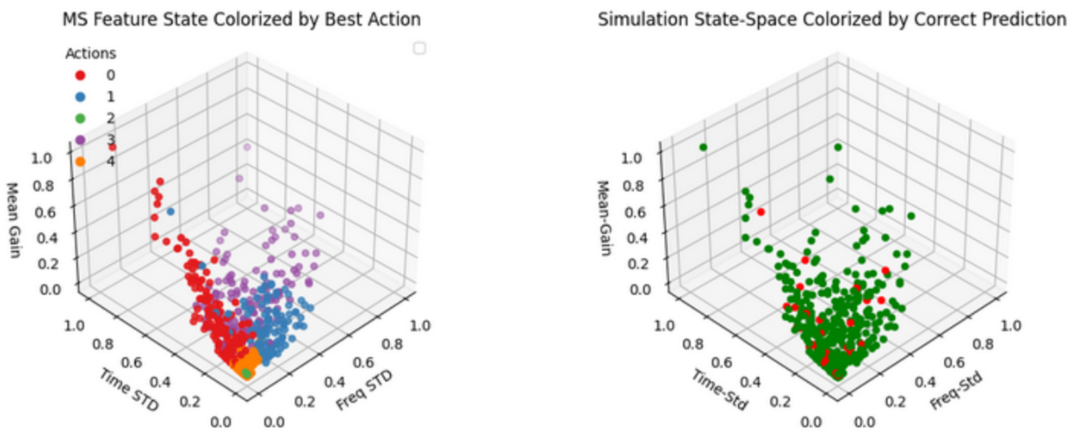
URSI GASS 2023, Sapporo, Japan, 19 – 26 August 2023

Artificially Intelligent Pipeline Sequencer

Brian M. Kirk*[1,2], Urvashi Rau [1], and Ramyaa[2]
(1) National Radio Astronomy Observatory, New Mexico, USA e-mail: bkirk@nrao.edu; rurvashi@nrao.edu
(2) New Mexico Institute of Mining and Technology, New Mexico, USA; e-mail: ramyaa.ramyaa@nmt.edu

Data taken with a radio interferometer must pass through a sequence of calibration and analysis stages in order to produce images that an astronomer can use to study astrophysical phenomena. Processing stages include data editing, calibration of the instrument hardware, and image reconstruction through deconvolution. Steps within each of these stages may be assembled in a variety of sequences and combinations, referred to as 'recipes'. Any single end-to-end data analysis recipe may be thought of as one possible path through all possible configurations. Overall, there is a predictable sequence of stages, but each stage may contain several optional paths depending on the characteristics of the outputs from the previous stage, and if certain conditions are not satisfied some stages (or steps) are revisited as subsequent steps, thus forming loops. In addition to sequencing stages and steps, most steps also need to decide parameters based on characteristics of the input data passed to that method. The full configuration space of the software tools is quite vast. Current automated pipelines essentially hard-code a "few", most-used paths through this entire maze, with a small set of heuristics-based decision points. Most pipeline executions still need to be reviewed by a trained Data Analyst for quality assurance and optimize individual steps as there is often no set of metrics to define correctness that applies in all situations.

Our research looks into developing a system that decides the optimal action based on the input data itself. This requires a software framework where actions can be evaluated and estimated within context. This research has started with building the fundamental components required to apply Deep Reinforcement Learning. This technique has the ability to evaluate possible actions with respect to global results and choose accordingly. When scaled, such methods are expected to learn the breadth of our high-dimensional, continuous, problem space and generalize the knowledge for inference. Our vision is that this generalized knowledge will be able to optimally calibrate and image datasets on its own.



Phase 1 results: These figures illustrate the state space spanned by simulations in a controlled environment designed to probe the cost-benefit of different data-averaging choices prior to calibration. The left panel shows ground-truth regions of which action is optimal for a dataset with the given characteristics. The right panel shows the accuracy achieved by the function mapping dataset characteristics to optimal action approximated by our neural network.