Debugging Bad Performance in Huge Infrastructure: Using ML and AI

Mariam Kiran, Anirban Mandal, George Papadimitriu, Cong Wag, Ewa Deelman
ESnet, RENCI, USC
mkiran@es.net, anirban@renci.org, georgpap@isi.edu, cwang@renci.org, deelman@isi.edu

Distributed systems present unique opportunities for massive computations and large data sets to be processed. However, the current state of the art uses these large infrastructures are often not optimally being used and there is little understanding to diagnose bad performances in very-large distributed jobs. In this paper, we argue that using AI and ML techniques, we can build "normal" profiles for these large infrastructures, that can help debug and improve the performance of the systems when they execute the distributed jobs.

INTRODUCTION

Scientific workflows involve analyzing, moving and processing large data files and simulations, arising from many experimental facilities, for example using the Large Hadron Collider (LHC) at CERN, climate modelling or the Advanced Light Source detector at Berkeley. With world-wide collaborations between research communities, network and computer engineers build and maintain a dedicated research infrastructure, dedicated to science needs. However, as these infrastructures grow, it becomes increasingly difficult to manage them, deduce if performance is degraded or providing improvements to their current performance. AI and ML has produced impressive results in automating a number of aspects such as ordering food, self-driving cars or even playing the game 'Go'. In this paper, we argue, that we can also use these techniques to improve how our infrastructure is being used, such as by building normal and abnormal profiles, trying better combinations to improve performance overall and providing recommended solutions to users on how the deploy their very large complex compute jobs.

In our previous work [1], we were able to show how engineers can deduce bad network throughput performances, computing errors and workflow run-time, by just monitoring the systems logs of the infrastructure while the jobs run on the system. In this work, we argue, that we can build on these experiences and build optimal performance profiles of the underlying structures to actually present improvements to future infrastructures.

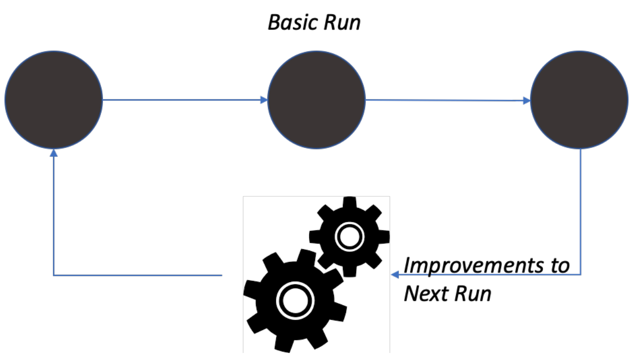

Figure 1: Improving performances from Basic Run.

1.1 Using AI/ ML Methods

In [2] we showed how machine learning can potentially provide more sophisticated and scalable methods to generate better performing workflows and possibly learn to create workflows based on prior knowledge. Figure 1 shows how previous runs of a workflow can be used to improve future runs. We anticipate the role of multiple machine learning techniques, in particular reinforcement learning to learn optimal combinations of the resources being used.

1.1.1 Reinforcement Learning.

Particularly, when models of the infrastructure are not available, we anticipate that reinforcement learning can allow model-free learning of optimal approaches to best optimize the current allocation of resources.

A reinforcement learning problem is formulated with an agent, situated in a partially observable environment, to learn from past data to make current decisions. The agent receives data in the form of snapshots of the environment, often with specific relevant features. Including information about value for future actions given the current state, the agent acts to change its environment, subsequently receiving feedback on the action in the form of rewards, (until terminal state is reached). The objective is to maximize the cumulative reward overall actions in the time the agent is active. Reinforcement learning research has investigated multiple techniques such as in multi-armed bandit problems, resource allocation or finding routes through a maze [3]. Deep reinforcement learning builds upon classical models, replacing the learning with a neural network to approximate policy and value functions. Here, the function approximates the environment state space with actions and rewards. Particularly when the state space is too large to store, this approach has proved feasible in learning approximate conditions.

1.2 Building Recommender Profiles

A recommender system can be used to provide the most relevant information to a user by discovering patterns in a dataset. We will develop an algorithm that rates the items and shows the user the items that they would rate highly. For workflow runs, building normal and abnormal profiles can help improve workflows performances in the future.

Conclusions

In the current state of art, workflow execution on large infrastructures lack the autonomic capability. This is a lack of methods that allows using spatial and temporal data dynamically from distributed facilities for model specification, calibration and validation. In our idea, using AI and ML we will bridge the gap between data observations and the specification of spatial models. In addition, this work proposes a conceptual framework for dynamic building and validation of workflow-driven computing devices for future large-scale infrastructures and science use cases.

REFERENCES

[1] M. Kiran, C. Wang, G. Papadimitriou, A. Mandal, and E. Deelman, “Detecting Anomalous Packets in Network Transfers: Investigations using PCA, Autoencoder and Isolation Forest in TCP,” Machine Learning, 2020.

[2] E. Deelman, A. Mandal, M. Jiang, and R. Sakellariou, “The role of machine learning in scientific workflows,” The International Journal of High Performance Computing Applications, 2019.

[3] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction. USA: A Bradford Book, 2018.