

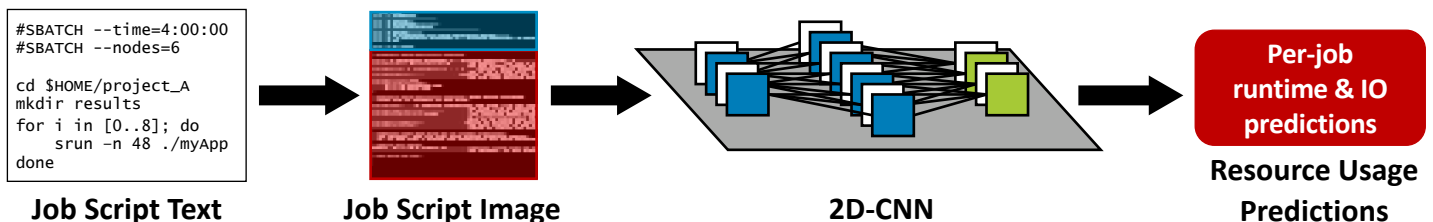
PRIONN

Predicting Runtime and IO using Neural Networks

M. Wyatt, S. Herbein, T. Gamblin, A. Moody, D. Ahn, M. Taufer

Project Overview

PRIONN (Predicting Runtime and IO using Neural Networks) is a tool that predicts resource usage of HPC jobs for IO-aware scheduling. User-submitted job scripts are mapped to image-like representations and used to predict resource usage of jobs using a 2D CNN. Resource usage predictions are used to enable resource-aware scheduling on next-generation HPC machines.

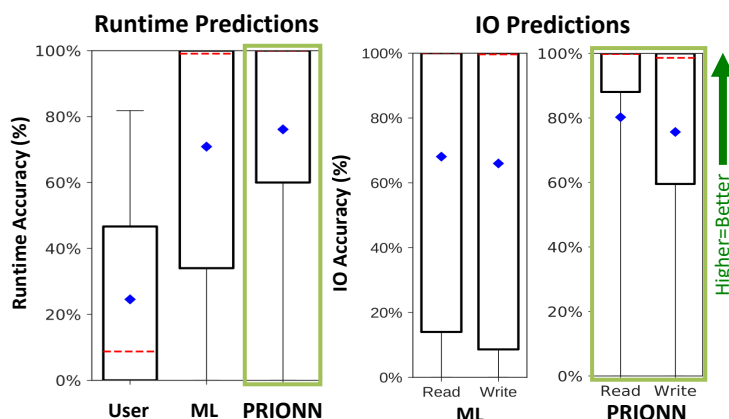
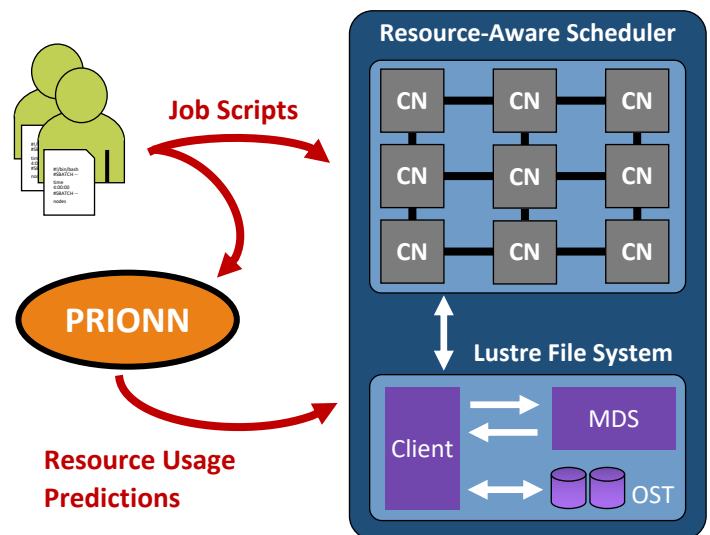


Motivation

Next generation HPC systems will need to be resource-aware to avoid resource contention (e.g., IO contention) among jobs. Resource contention is a significant contributor to interference between jobs that can slow execution. Resource-aware scheduling is one way that resource contention can be avoided. PRIONN is designed to run alongside HPC schedulers to provide the runtime and resource usage predictions needed for a resource-aware scheduler.

PRIONN:

- Requires no a priori job script processing
- Is compatible with any job script
- Provides accurate resource usage predictions



Results

PRIONN is more accurate at predicting resource usage (e.g., runtime and IO) than users and traditional machine learning approaches:

- >75% mean runtime accuracy
- >75% mean IO read/write accuracy
- >55% of future system IO bursts predicted

Publication

M. Wyatt, S. Herbein, T. Gamblin, A. Moody, D. H. Ahn, and M. Taufer: **PRIONN: Predicting Runtime and IO using Neural Networks**, ICPP 2018: 46:1-46:12.

