Report on Ensemble Prediction Slurm

Introduction:

This paper proposes a machine learning (ML)-based method to accurately predict the memory and time requirements of jobs submitted to High Performance Computing (HPC) systems using SLURM(Simple Linux utility for Resource Management), a popular job scheduler. This helps avoid overestimation by users, which can waste resources and lower system efficiency.

Important Terminologies:

1). SLURM = SLURM stands for Simple Linux Utility for Resource Management. Slurm is an open source, fault-tolerant, and highly scalable cluster management and job scheduling system for large and small Linux clusters. Slurm requires no kernel modifications for its operation and is relatively self-contained. As a cluster workload manager, Slurm has three key functions.

First, it allocates exclusive and/or non-exclusive access to resources (compute nodes) to users for some duration of time so they can perform work. Second, it provides a framework for starting, executing, and monitoring work (normally a parallel job) on the set of allocated nodes.

Finally, it arbitrates contention for resources by managing a queue of pending work.

2). XSEDE = XSEDE (Extreme Science and Engineering Discovery Environment) is a distributed network of high-performance computing resources, data storage, and related services for researchers. XSEDE Service Providers (SPs) are the institutions that operate and manage these resources.

They provide access to computational resources, storage, and other services that researchers need for their work.

3). BEOCAT = Beocat is the High-Performance Computing (HPC) cluster operated by the Institute for Computational Research at Kansas State University.

One of the largest academic supercomputers in Kansas, supporting nearly 400 researcher-funded machines, ~10,000 CPU cores, over 3.3 PB of storage, and around 170 GPUs, including NVIDIA L40S nodes.

4). Sun Grid Engine = Sun Grid Engine (SGE), later acquired and renamed by Oracle as Oracle Grid Engine, is a software system for managing and distributing workloads across a cluster of computers. It acts as a batch scheduler, allowing users to submit jobs to the cluster and have them automatically executed on available resources.

SGE is designed to handle large computational tasks, such as simulations or video conversions, by distributing them across multiple machines.

5). Tera-Scale Open Source Resource and Que Manager(TORQUE) = TORQUE, is a distributed resource manager designed to oversee batch jobs and distributed compute nodes. It provides control and management capabilities for clusters, aiding in utilization, scheduling, and administration tasks.

Originally based on the Portable Batch System (PBS), TORQUE has evolved with significant contributions from various HPC entities, improving scalability, fault tolerance, and overall functionality.

6). Portable Batch System (PBS) = The PBS is a job scheduling and workload management system designed to manage and distribute computing tasks (jobs) across the nodes of a High Performance Computing (HPC) cluster or supercomputer. It enables batch processing, job queuing, resource management, and scheduling.

7). Mauri Cluster Scheduler = Mauri cluster is an HPC system managed by the Institute for Geosciences and Petroleum (IGP) at the Norwegian University of Science and Technology. It is based on Linux. It supports computationally-intensive tasks in petrology, geophysics, mineral physics, and related research.

It is a GPU accelerated Linux compute platform.

Concepts:

1). CPU time raw: It is a metric used in Slurm (Simple Linux Utility for Resource Management) that represents the total CPU time consumed by a job.

Mathematically, it is calculated as:

CPU-Time-RAW = Elapsed time x number of CPU used

Here,

Elapsed time = The wall-clock time from when the job started running to when it finished.

Number of CPUs = number of processor cores allocated to the job.

2). Linear Regression: Linear Regression analysis is a powerful technique used for predicting the unknown value of a variable(Dependent variable) from the known value of another variables(Independent variable).

A dependent variable is the variable to be predicted or explained in a regression model.

An independent variable is the variable related to the dependent variable in a regression equation.

3). Max RSS: Max-RSS stands for Maximum Resident Set Size, and in the context of Slurm and HPC job accounting, it represents the maximum amount of physical memory (RAM) used by a job (or all of its tasks) at any one time, measured in kilobytes (KB) or megabytes (MB). It tells you how much memory your job actually needed while running. It also helps users and administrators avoid over-requesting or under-requesting memory.

4). Ridge Regression: It is a technique used in linear regression to address multi collinearity and prevent overfitting by adding a penalty term to the cost function.

This penalty, known as L2 regularization, shrinks the coefficients towards zero, but unlike Lasso regression, it doesn't force them to zero, meaning it doesn't perform feature selection.

5). Lasso Regression: Performs both regularization and feature selection. It is a linear regression technique that performs both variable selection and regularization. It's used to improve model accuracy and interpretability, especially when dealing with high-dimensional data or multi collinearity.

6). Decision Trees: A decision tree is tree like structure which mimics the process of decision making by breaking down data into smaller subsets based on specific features.

Each internal node of the tree represents a decision based on a feature, and each leaf node represents the final prediction or decision.

7). Random Forest Regression: Random forest is used for both classification and regression tasks. It is an ensemble method that combines multiple decision trees to make predictions.

Instead of relying on a single decision tree, it creates a forest of multiple decision trees. Each tree is trained on a random subset of the data, by introducing randomness, each tree becomes slightly different from others.

8). LightGBM: Light gradient boosting machine is an open-source high-performance framework developed by Microsoft. It is an ensemble learning framework that uses gradient boosting method which constructs a strong learner by sequentially adding weak learners in a gradient descent manner.

It's designed for efficiency, scalability and high accuracy particularly with large datasets. It uses decision trees that grow efficiently by minimizing memory usage and optimizing training time. Key innovations like Gradient-based One-Side Sampling (GOSS), histogram-based algorithms and leaf-wise tree growth enable LightGBM to outperform other frameworks in both speed and accuracy.

9). Elastic net regression: Elastic-Net Regression is a modification of Linear Regression which shares the same hypothetical function for prediction.

It is particularly useful when dealing with datasets that have a large number of features, especially when some features are highly correlated.

10). Scikit learn: Scikit-Learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is probably the most useful library for machine learning in Python.

The library is built upon the scipy that must be installed before you can use scikit-learn.

Methodology:

1). Data preparation: Two datasets from the RMACC Summit and Beocat clusters (spanning millions of job entries) were collected. Each job log included features like requested memory, CPU count, job duration, QOS, and others. The authors cleaned and filtered the data by removing incomplete jobs or anomalies. They then normalized and standardized the features to prepare them for model training. The main prediction targets were cpu time raw (a composite runtime measure) and max-rss (memory usage).

2). Machine Learning Algorithms: Seven ML regression algorithms were evaluated: Lasso, Linear Regression, Ridge, Elastic Net, Decision Trees (DTR), Random Forest (RFR), and LightGBM (LGBM).

These models were trained to predict the CPUTimeRAW and MaxRSS values using the prepared dataset. The authors used R^2 and RMSE as metrics to evaluate model performance. Scikit-learn and LightGBM libraries were used for implementation.

3). Mixed Account Regression Models (MARM): Instead of training on all job data at once, the authors grouped jobs by user account and trained models on subsets. This approach, called Mixed Account Regression Model (MARM), improves prediction accuracy by leveraging the fact that users often submit similar types of jobs.

The algorithm adds accounts incrementally to the model based on how much they improve R^2 score, optimizing performance on smaller data partitions while still generalizing well.

Results Obtained:

1). Up to 86% in predicting memory and time.

2). For RMACC:

- Waiting time (WT) is reduced from 380 to 4 hours.

- Turnaround time (TAT) is reduced from 403 to 6 hours.

- Utilization is improved to 100%.

3). For Beocat:

- Waiting time (WT) is reduced from 662 to 28 hours.
- Turnaround time (TAT) is reduced from 673 to 35 hours.
- Utilization is achieved 100%. Same as RMACC.