Anomalies Diagnosis in High Performance Computing using Machine Learning

Muna Tageldin

November 16, 2022

With the rapid growth of High-performance computing technologies and their complexity, predicting the performance of HPC applications at scale is a well-known problem. Complex system architectures, such as memory hierarchies, or interconnection between the application and system software/hardware under different architectures, resource requirements, etc present significant challenges to application performance analysis and prediction. These interactions are manifested in a system error, memory access, storage, etc causing anomalous behavior at scaled HPC applications. The anomalous behavior experienced in these systems causes major slowdowns, loss of computing resources, or inefficient scheduling. Therefore, designing and developing efficient anomalous detection systems are vital for the efficiency of HPC applications. Machine learning algorithms are an active field in this research area.

0.1 anomalous detection model

The proposed machine learning model for anomalous detection comprises 1) a feature extraction unit and 2) a machine learning model 3) performance metrics.

0.1.1 Feature Extraction

The data used for the anomaly detection system are collected from HPC performance analysis reports. These data are usually time series data. Wavelet Transform along statistical features like Kurt and skew are extracted from the collected data. Wavelet Transform provides an efficient methodology to learn dynamic patterns in a sequence by convolving the time domain signal with a wavelet family function. Kurt and skew are statistical measures that measure the symmetry and heavy-tailness of the kernel density distribution estimated from the data, respectively [3], [4].

0.1.2 machine learning algorithm

An anomaly detection system could be developed by using an autoencoder. Autoencoders are a special type of Neural network that learn pattern representations from data. Specifically, an autoencoder encodes the input or maps

the high-dimensional input to a low-dimensional output. Then, the autoencoder decoder constructs the input from the encoder output. This process of encoding and decoding allows the discovery of hidden patterns and relationships in the data hence, detecting anomalous events. Specifically, if the reconstructed input does not match the original input, an anomalous event is detected [1] , [2].

0.1.3 Models Training and Performance Metrics

The features collected are divided into 1)70% to train the classifier and 2) 30% to validate the proposed classifier. The confusion matrix and miss rate are used to measure the performance of the proposed anomaly detection algorithm.

References

[1] Burak Aksar et al. "Proctor: A Semi-Supervised Performance Anomaly Diagnosis Framework for Production HPC Systems". In: *High Performance Computing*. Ed. by Bradford L. Chamberlain et al. Cham: Springer International Publishing, 2021, pp. 195–214. ISBN: 978-3-030-78713-4.

[2] Burak Aksar et al. "Proctor: A Semi-Supervised Performance Anomaly Diagnosis Framework for Production HPC Systems". In: *High Performance Computing*. Ed. by Bradford L. Chamberlain et al. Cham: Springer International Publishing, 2021, pp. 195–214. ISBN: 978-3-030-78713-4.

[3] Ozan Tuncer et al. "Online Diagnosis of Performance Variation in HPC Systems Using Machine Learning". In: *IEEE Transactions on Parallel and Distributed Systems* 30.4 (2019), pp. 883–896. DOI: 10.1109/TPDS.2018.2870403.

[4] Yueyue Yao, Jianghong Ma, and Yunming Ye. "Regularizing autoencoders with wavelet transform for sequence anomaly detection". In: *Pattern Recognition* 134 (2023), p. 109084. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2022.109084. URL: https://www.sciencedirect.com/science/article/pii/S0031320322005647.