Desh: Deep Learning for HPC System Health Resilience

Anwesha Das⁠¹, Abhinav Vishnu⁠², Charles Siegel⁠², Frank Mueller⁠¹
¹North Carolina State University  ²Pacific Northwest National Lab

Motivation

- **Problem**: Challenges - Can failed event truly indicate failure? How to distinguish real failures from noise and benign events? Is a scalable automated framework possible?
- **Goal**: Investigate deep learning techniques such as LSTM for HPC system failure prediction. Research methods to scale training phase of logs and predict sensible events.

Solution Paradigm

- **Text Phrases from Logs**: Expert guided filters
- **Phrase Grouping**: Deep learning: Efficient in vision and speech recognition. Can it enhance reliability of HPC systems?
- **Required**: Efficient scalable solutions for failure prediction, proactive recovery

- **System Details**: System: Cray XC Cluster1 (C1) Data Size: 20 MB Duration: 1 week

Desh Prototype

- **Why LSTM?**: Can unlearn and relearn time-series data
- **Ability to capture long-term and short-term data correlations**
- **Known to be efficient for large scale data processing**
- **To understand whether Neural Networks unlike Markov Models can predict derived phrases learned from the training data.**

Background

- **Past Research**: Anomaly detection/prediction for older HPC systems
- **Past Logs**: Comparatively more structured
- **Past Focus**: Statistical Analysis, Inadequate stress on text semantics & lead times
- **Contemporary HPC systems**: New format, unstructured text logs
- **New scope**: Natural Language Processing (NLP), Deep Learning [3] based Techniques
- **Past Techniques**: Logistic regression, PCA (principle component analysis) [4], Event correlation, Probabilistic Model and Markov Chain based mechanisms
- **Future Work**: Learn feature extraction: Supervised or easier to do labeling
- **Support Vector Machines (SVMs)** [1] & Sequence Mining [2] based mechanisms
- **Conclusion**: Identified scopes to improve HPC system health considering phrase embeddings and semantics for better lead times.

Results

- **FPR (False Positive Rate)**: Phrases which didn’t appear in the test data, but Desh predicted, (depends on training set)
- **Lead Time**: The correctly predicted phrases are cross validated in the data, to know how much ahead in time, the phrases actually occur (after the last trained phrase)

Future Work

- **How little expert labeling can auto-classify the predicted phrases?**
- **How to analyze the unknown class for understanding which phrases are mostly safe or part of an anomaly?**
- **How to predict future time-series accurately to aid failure prediction with location information?**
- **Comparative analysis of Desh with existing prediction techniques on multiple HPC cluster logs.**

Acknowledgments: Dr. Abhinav Vishnu, Dr. Charles Siegel and Dr. Frank Mueller for insightful guidance and helpful suggestions. The CIF and EMSL division of PNNL for cooperating with HPC cluster access and data sharing.