

# 21st International Conference on Computing in High Energy and Nuclear Physics (CHEP2015)



Contribution ID: 211

Type: poster presentation

## Applying deep neural networks to HEP job statistics

The cluster of CC-IHEP is a middle sized computing system providing 10 thousands CPU cores, 3 PB disk storage, and 40 GB /s IO throughput. Its 1000+ users come from serials of HEP experiments including ATLAS, BESIII, CMS, DYB, JUNO, YBJ etc. In such a system, job statistics is necessary to find performance bottlenecks, locate software pitfalls, identify suspicious behaviors and make resource provisions, especially for new experiments whose computing modeling are still developing and refining.

To fulfill this requirement, we have developed and deployed a job statistics system which consists of an instrumenting agent, a central database, a data summarizer and a visualizer on the IHEP cluster. In the first half of 2014, the system has collected 1 million valid job records from BESIII experiment. Each job record includes static information from batch system, average efficiency from process manager and detailed IO parameters from VFS interfaces. To analyze this dataset we find that DNNs (Deep Neural Networks) is a useful technique for data classification and abnormality detection.

This paper demonstrates how we train a job classifier with DNNs. It firstly describes how we label the dataset semi-automatically from about 20% jobs samples which have hints of job type in their job-option-file names. Then some adapted data pre-processing steps will be presented. After that, it will describe the DNNs model which has achieved a precision of 96.6% with 240'000 labeled job samples (Ratio of training set and testing set is 7:3), 6 classes. It will also compare the results with those from a linear model and a MLPs (Multi-Layer Perceptrons) model. Impacts of meta-parameters including learning rate, batch size will be discussed. Examples of how we leverage the classification results to find software problems and detect abnormal job behaviors will be given at the last part of this paper.

**Primary author:** WANG, Lu

**Presenter:** WANG, Lu

**Track Classification:** Track6: Facilities, Infrastructure, Network