Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st André Lucena Ribas Ferreira
University of Minho
Braga, Portugal
pg52672@uminho.pt

1st Carlos Eduardo da Silva Machado
University of Minho
Braga, Portugal
pg52675@uminho.pt

1st Goncalo Manuel Maia de Sousa
University of Minho
Braga, Portugal
pg52682@uminho.pt

***Abstract*—This document is a model and instructions for LaTeX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

- Machine learning has increased in popularity
  - image classification
  - natrual language proccessing
- studies have tried to analyse I/O patterns in DL Workflows (source)
- very few get down to kernel level
- eBPF are ...
- we seek to provide a tool to Characterize DL workloads using eBPF's

II. Background

- DL involves iterating multiple times (epochs) through a dataset
- all data is read exactly once one epoch as passed (I/O intensive)
- passing it through all the layers to calculate a loss (forward pass)
- use calculated loss to update the learnable parameters of the network (backpropagation)
- SGD is an optimizer for loss function minimization widely used for its lower computation
- DL is usually I/O-bound [need source], due to the use of accelerators (GPU), size of the data and random reads
- pytorch is a DL framework
- Distributed DNN training (data paralellism)
- checkpointing involves saving the model state
- in pytorch its done explicity with torch.save() and in official workloads is done in-between epochs
- eBPF's

III. Related work

- papers que usam darshan/tf-darshan para caracterizar padrões
- MLPerf Storage/tese de um aluno da Oana
- DIO e tools de observabilidade que usam eBPF e outras (related work do DIO), LD PRELOAD, captura de de I/O request por intrumentação do código fonte.
- Caracterizar Tensorflow I/O workloads através da sua definição teórica [1]
- O que falta fazer?
  - Análise empírica dos padrões como parte do processo de treino
  - Análise da cache como interveniente no processo de I/O
  - Testes de Rede (para modelos distribuídos)
  - Analisar PyTorch

IV. Design

- Grafana
- python parser and plots

V. Evaluation Methodology

- dstat, nvidia-smi to get cost of using the tool
- grafana dashboard to get data

VI. Evaluation Results

VII. Conclusion

References

[1] S. W. D. Chien, S. Markidis, C. P. Sishtla, L. Santos, P. Herman, S. Narasimhamurthy, and E. Laure, "Characterizing deep-learning i/o workloads in tensorflow," in *2018 IEEE/ACM 3rd International Workshop on Parallel Data Storage & Data Intensive Scalable Computing Systems (PDSW-DISCS)*, 2018, pp. 54–63.