ACAT 2022

Sunday, 23 October 2022 - Friday, 28 October 2022

Villa Romanazzi Carducci, Bari, Italy

Book of Abstracts

values. This scenario will also imply an unprecedented increase of simulation and collision data to transfer, process and store in disk and tape systems. The Spanish WLCG sites that support CMS, the PIC Tier-1 and the CIEMAT Tier-2 have explored content delivery network type solutions in the Spanish region. One of the possible solutions under development has been the deployment of caches between the two sites that store the data requested by the jobs remotely, so that they get closer to the nodes to improve their job efficiency and input data transfer latency. In this contribution, we analyze the impact of deploying physical caches in production in the CMS region between PIC and CIEMAT, as well as the impact they have on job efficiency, latency and bandwidth gains, and potential storage savings.

Significance:

The relevance of this contribution is to discuss the results of the new data management system in the Spanish CMS region in convergence with the WLCG-DOMA data challenges.

References:

1 J. Albrecht, et al, "A Roadmap for HEP Software and Computing RD for the 2020s", Computing and Software for Big Science volume 3, Article number: 7 (2019) https://doi.org/10.1007/s41781-018-0018-8.
2 CMS data access and usage studies at PIC Tier-1 and CIEMAT Tier-2,
Antonio Delgado Peris, José Flix Molina, José M. Hernández, Antonio Pérez-Calero Yzquierdo, Carlos Pérez Dengra, Elena Planas, Francisco Javier Rodríguez Calonge, Anna Sikora, EPJ Web Conf. 245 04028 (2020).
3 Pérez Dengra, C., 2022 "Simulating a network delivery content solution for the CMS experiment in the Spanish" WLCG Tiers, International Symposium on Grids & Clouds (ISGC) 2022 Virtual Conference, 21-25 March 2022, last access on 25th of May of 2022: https://indico4.twgrid.org/event/20/contributions/1116/

Experiment context, if any:

CMS Collaboration

Track 3: Computations in Theoretical Physics: Techniques and Methods / 173

Product Jacobi-Theta Boltzmann machines with score matching

Authors: Andrea Pasquale[1]; Stefano Carrazza[2]

[1] *University of Milan*

[2] *CERN*

Corresponding Author: andrea.pasquale@unimi.it

We introduce a restricted version of the Riemann-Theta Boltzmann machine, a generalization of the Boltzmann machine with continuous visible and discrete integer valued hidden states. Though the normalizing higher dimensional Riemann-Theta function does not factorize, the restricted version can be trained efficiently with the method of score matching, which is based on the Fisher divergence. At hand of several common two dimensional datasets, we show that the quality of the fits obtained are comparable to state-of-the-art density estimation techniques such as normalizing flows or kernel density estimation. We also discuss how some of these methods can converge to an overfitted solution and we try to quantify this overfitting behavior.
Furthermore, we show that our model is less likely to converge to such non ideal solutions.
We also prove that the recursive calculation of the one dimensional Riemann-Theta function can be extended to the calculation of the first and second order gradients.
We also hint at the possibility of using the density estimated by this model
to perform multi-dimensional integration using Monte Carlo methods with a particular focus on High Energy Physics applications.

Significance:

The major updates on the Riemann-Theta Boltzmann machine are the possibility to train efficiently systems with more than 2 hidden layer thanks to the method of score matching. We also present a novel way to
quantify the overfitting through the surface fractal dimension.

References:

https://inspirehep.net/literature/1644620
https://inspirehep.net/literature/1694236
https://inspirehep.net/literature/1737266

Experiment context, if any:

Poster session with coffee break / 174

Optimizing electron and photon reconstruction using deep learning: application to the CMS electromagnetic calorimeter

Author: Davide Valsecchi[1]

[1] *ETH Zurich (CH)*

Corresponding Author: davide.valsecchi@cern.ch

The reconstruction of electrons and photons in CMS depends on topological clustering of the energy deposited by an incident particle in different crystals of the electromagnetic calorimeter (ECAL). These clusters are formed by aggregating neighbouring crystals according to the expected topology of an electromagnetic shower in the ECAL. The presence of upstream material (beampipe, tracker and support structures) causes electrons and photons to start showering before reaching the calorimeter. This effect, combined with the 3.8T CMS magnetic field, leads to energy being spread in several clusters around the primary one. It is essential to recover the energy contained in these satellite clusters in order to achieve the best possible energy resolution for physics analyses.

Historically satellite clusters have been associated to the primary cluster using a purely topological algorithm which does not attempt to remove spurious energy deposits from additional pileup interactions (PU). The performance of this algorithm is expected to degrade during LHC Run 3 (2022+) because of the larger average PU levels and the increasing levels of noise due to the ageing of the ECAL detector. New methods are being investigated that exploit state-of-the-art deep learning architectures like Graph Neural Networks (GNN) and self-attention algorithms. These more sophisticated models improve the energy collection and are more resilient to PU and noise.

This contribution covers the model optimization results and the steps to put it in production inside the realistic CMS reconstruction sequence. The impact on the electron and photon energy resolution and tests of the resiliency of the algorithm to the changing detector conditions are shown.

Significance:

The topic of this talk has been shown for the first time at ACAT2021. This new contribution
will cover the process to put the ML algorithm into production in the CMS software and the optimization of the model to obtain the final physics performance for electron and photon reconstruction and the best computational efficiency.

References:

https://arxiv.org/pdf/2204.10277.pdf

Experiment context, if any:

CMS