



Contribution ID: 42

Type: YSF poster

Machine Learning based Compression of Scientific Data - the HEP Perspective

Monday 30 October 2023 18:35 (20 minutes)

One common issue in both research and industry is the growing data volumes and thereby the ever-increasing need for more data storage. With experiments taking more complex data at higher rates, the data recorded is quickly outgrowing the storage capabilities [1]. Since the data formats used are already highly compressed, storage constraints would require more drastic measures such as more exclusive event selection where a large portion of the data is discarded, or lossy compression, where data can be compressed beyond traditional lossless techniques as a result of some loss in resolution.

As a potential solution to tailored lossy compression, we present Baler - an interdisciplinary, open-source, open-access tool for machine learning-based data compression. The tool uses autoencoders trained to compress and decompress data based on learned correlations. Interesting caveats are presented between offline and online compression, with studies on ways to efficiently overfit the data in the former. We show that, for common observables in high energy physics, where the precision loss is tolerable, the high compression ratio allows for more data to be stored yielding greater statistical power.

[1] - <https://cerncourier.com/a/time-to-adapt-for-big-data/>

Brainstorming idea [title]

Zero-Knowledge Machine Learning based Anomaly Detection in High Energy Physics

Brainstorming idea [abstract]

Zero knowledge proofs are a concept that originate from cryptography, where a prover (algorithm/agent) proves that a statement is true to a verifier (algorithm/agent) without revealing any further information about the statement. This concept has recently been applied to multiple areas of machine learning greatly improving inference speeds by reducing the net number of computations required to, say classify an instance to a fixed number of classes.

There are multiple other extensions of ZKP into ML as well and a comprehensive, promptly updated repository for all things ZKML is available here: <https://github.com/zkml-community/awesome-zkml>

I would like to discuss the potential uses of ZKML across various problems in HEP, with my personal interests lying in anomaly detection for new physics searches. I am open to start a discussion on other potential applications as well as extensions of the ZKML area of research.

Primary authors: GALLEN, Axel (Uppsala University (SE)); DOGLIONI, Caterina (University of Manchester (GB)); EKMAN, Per Alexander (Lund University (SE)); JAWAHAR, Pratik (University of Manchester (UK - ATLAS))

Presenter: JAWAHAR, Pratik (University of Manchester (UK - ATLAS))

Session Classification: Young Scientist Forum