Abstract for MSc Data Science Thesis:
Machine Learning for Particle Identification and Deep Generative Models for High Energy Physics Detector Simulations
(Christiaan Gerhardus Viljoen, VLJCHO04)

Abstract

This Masters project was focused on the application of machine learning techniques towards specific aspects of particle physics. Its two main aims: *particle identification* and *high energy physics detector simulations* are pertinent to research avenues pursued by physicists working with the ALICE[1] TRD[2] detector, within the LHC[3] at CERN[4].

Aims

More formally, the aims of this project were as follows:

1. For particle identification: various neural networks were trained and assessed, to determine their ability to discriminate between electrons and pions, produced during proton-Lead (pPb) collisions conducted at the LHC in 2016, based on ADC[5] signal data produced as these particles were detected by the ALICE TRD. (Note that this work was done on uncalibrated raw TRD digits).
2. For high energy physics detector simulations: Geant4, a Monte Carlo toolkit used to simulate particle interactions with matter, was assessed in terms of how closely the simulated data it produces resembles true data taken by the TRD during collision events. In addition, as a step towards fast simulation, various deep generative modeling strategies were employed to produce simulated data samples which are likely under the observed (true) TRD data distribution. To this end, the following classes of latent variable models were prototyped: Generative Adversarial Networks, Variational Autoencoders and Adversarial Autoencoders. Data produced during these deep generative simulations were compared to real data in the same manner as that done for Geant4 data, in order to assess the feasibility of incorporating these types of models into future high energy physics event simulation software.

Summary of Results

Particle identification performance was defined by the ability of each neural network to minimize pion efficiency (ε_π, false positive rate), whilst maximizing electron efficiency (ε_e, true positive rate). A lower bound for the critical region (t_{cut}) in the distribution of $P(elec)$ predictions made by each neural network which results in $\varepsilon_e \approx 90\%$ was defined, in order to determine the ε_π for that neural network. The best set of results obtained, per momentum bin, was as follows: $\varepsilon_\pi = 1.2\%$ in the $p \leq 2$ GeV/c range; $\varepsilon_\pi = 1.14\%$ in the 2 GeV/$c < p \leq 3$ GeV/c range; and $\varepsilon_\pi = 1.51\%$ in the 3 GeV/$c < p \leq 4$ GeV/c range.

[1] A Large Ion Collider Experiment

[2] Transition Radiation Detector

[3] Large Hadron Collider

[4] European Organization for Nuclear Research

[5] Analog to Digital Converter

Abstract for MSc Data Science Thesis:
Machine Learning for Particle Identification and Deep Generative Models for High Energy Physics Detector Simulations
(Christiaan Gerhardus Viljoen, VLJCHO04)

In terms of results obtained for high energy physics detector simulations, distinguishing Geant4 data from real data was a trivial task when compared to the task of particle identification. Similarly, data produced by deep generative models were easily distinguishable from real data; but the obtained results (especially for variational- and adversarial autoencoders) appear to be promising enough to pursue in future research.

Keywords

Deep Learning, Convolutional Neural Networks, Particle Identification, High Energy Physics Detector Simulations, Generative Adversarial Networks, Variational Autoencoders, Adversarial Autoencoders, Geant4, ROOT, AliROOT, Tensorflow, Keras