

Christiaan Gerhardus Viljoen – MSc Data Science – University of Cape Town – Accepted January 2020

Abstract

This Masters thesis outlines the application of machine learning techniques, predominantly deep learning techniques, towards certain 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 (A Large Ion Collider Experiment) Transition Radiation Detector (TRD), within the Large Hadron Collider (LHC) at CERN (The European Organization for Nuclear Research).

Aim 1: Particle Identification

The first aim of this project focused on the application of machine learning techniques towards particle identification; in particular, the classification of electrons (e) versus pions (π) produced during proton-Lead (pPb) collisions during various runs from LHC16q. Various neural network architectures, hyperparameter settings, etc. were assessed by optimising an electron acceptance cut-off point (t_{cut}) in the distribution of the classifying neural network's $P(electron)$ estimates, which minimises the amount of pion contamination (i.e. pion efficiency, ε_π), whilst maintaining a high rate of electron acceptance (i.e. electron efficiency, ε_e), specifically $\varepsilon_e \approx 90\%$.

Summary of Results for Particle Identification

Particle identification in this thesis was performed on uncalibrated TRD digits data, in contrast to work that has been done in this area before. For this reason, the presented results are only comparable, in terms of accuracy, to some of the less useful methods that have been investigated by others. Nonetheless, a much more comprehensive search across the space of possible neural network types and architectures was done in this project and this exploratory work has resulted in some interesting findings coming to light, such as the usefulness of the Focal Loss function in mitigating for the extreme class imbalances present in this dataset. The main goals of the first aim of this thesis were: (1) to show that the obtained results were comparable to previous studies, with slightly inferior performance (as expected, due to the omission of the calibration phase of data pre-processing); and 2) getting to know the dataset at hand well enough to enable the second aim of this thesis, which is discussed below.

Aim2: High Energy Physics Detector Simulations

The second aim of this project centred around determining whether detector simulations obtained from Geant4 were as accurate as they are usually assumed to be. Additionally, for the first time, exploratory research was done into the feasibility of making use of latent variable/ deep generative models for fast detector simulations for the ALICE Transition Radiation Detector. To this end, a wide variety of generative models were prototyped, and the results of a few choice models will be presented.

Summary of Results for High Energy Physics Detector Simulations

Perhaps the most important result of this thesis is the fact that a convolutional neural network was able to distinguish between simulated data (obtained by making use of Geant4 as the detector simulation component) and true data obtained by the TRD during a specific LHC pPb run. An investigation is presented which delves into some of the major differences between the two types of data.

Additionally, various deep generative models were prototyped towards the simulation of the TRD detector response, some of these models produced results which could indicate that these models would be fruitful avenues to explore as part of the detector simulation component of the O² software currently being developed for LHC Run 3.