Vrije Universiteit Amsterdam

Universiteit van Amsterdam

Master Thesis

Title of the Thesis

Author: student name (student number)

1st supervisor:	supervisor name	
daily supervisor:	supervisor name	(company, if applicable)
2nd reader:	supervisor name	

A thesis submitted in fulfillment of the requirements for the joint UvA-VU Master of Science degree in Computer Science

July 28, 2020

"I am the master of my fate, I am the captain of my soul"
from Invictus*, by William Ernest Henley*

Abstract

Here goes the abstract of this thesis.

Contents

List of Figures

LIST OF FIGURES

List of Tables

1

Introduction

The KM3NeT or the Cubic Kilometer Neutrino Telescope is currently being constructed at the bottom of the Mediterranean Sea. The goal of this telescope is two fold: first is to study high energy neutrinos originating from celestial events such as birth of a neutrino star, supernovae, etc. and second, to study the properties of the neutrino particles produced in the Earth's atmosphere (1). The first goal will be realized with the KM3NeT/ARCA (Astroparticle Research with Cosmics in the Abyss) telescope and the second with KM3NeT/ORCA (Oscillation Research with Cosmics in the Abyss) (1). In this paper, we talk exclusively about KM3NeT/ARCA.

The ARCA telescope comprises of two "blocks" with a total volume of $1km^3$. Each block consists of 115 spherical detector units (referred to as DOMs henceforth) and each DOM consists of 31 Photo Multiplier Tubes (PMTs) in various spacial arrangement. Figure **??** shows an artist's impression of ARCA, figure **??** depicts a DOM along with the PMTs inside it.

1.1 Context

The PMTs are sensitive to light or photons, the analog signal for all hits above a certain threshold are digitized. This datapoint consists of a timestamp and the spatial orientation of the DOM (ie. x,y,z coordinates). The digital signals from all PMTs are arranged in 100ms "time slices" and sent to the on-shore facility for further processing (2).

1.2 Situation of Concern

When the high energy neutrino particles interact with surrounding matter, produce an electron and a photon, this phenomenon is known as the Cherenkov Radiation or Cherenkov

Light (3). This phenomenon is utilized in the KM3NeT telescopes to detect high energy neutrinos ie. the Cherenkov Light is detected by the PMTs. Unfortunately, there are several sources of noise (in this case, the noise is other light sources), bioluminescense and decay of Potassium 40 (^{40}K) and atmospheric Muons being the primary sources (4).

1.2.1 State of the Art

Due to the high level of noise, data is generated at an extremely high rate of 25GB/sec (1). Due to this high data rate, it must be filtered and selectively stored for further analysis. The state of the art for this task are known as "Event Trigger" algorithms (1, 2). The existing event trigger algorithms namely *L1* and *L2* have limitations and can be improved upon (5).

There is a tradeoff between performance and quality of the event trigger algorithms to be realized here. The state of the art *L1* and *L2* algorithms are very performant with the ability to filter data in real time however their quality of filtration can be improved. Thus new event trigger algorithms must consider this trade off.

Efforts have already been made to improve the existing event trigger algorithms. (5) proposed and implemented a GPU powered pipeline which utilizes correlation and graph community detection to identify time slices that may contain neutrino hits whilst (4) suggests an alternate using convolutional neural networks.

1.2.2 User Requirements

The primary users of the ARCA are researchers who want to study high energy particles from outer space. The stakeholders are all member institutes involved in the project and by extension all scientists from these institutes who will be working with the data collected.

The requirements of the primary users (and stakeholders) with respect to the data acquisition pipeline are as follows: 1. The accuracy with which time slices are filtered out has to be extremely high. Time slices which are deemed important by event trigger algorithms are stored for further analysis and research. Failure to store time slices containing important data can lead to loss of important data and thus a poor quality of research. 2. The filtration of noise must be highly accurate. The event trigger algorithms must be able to eliminate majority of noise to prevent storage of unnecessary and potentially useless data. 3. The state of the art event trigger algorithms are able to process data in real time. The proposed alternative ideally should maintain or improve upon it's predecessor's performance else provide a good trade off with data quality.

1.3 Research Question

This report intends to improve upon the GPU pipeline proposed by (5), specifically we wish to answer the following research questions:

1. RQ1 Can the existing GPU pipeline be improved using neural networks (NNs)?
2. RQ2 Which NN will produce the best results when replaced with the *Pattern Matrix*?
3. RQ3 Which NN will produce the best results when replaced with the *Graph Community Detection* algorithm?
4. RQ4 Finally, which combination of NNs will produce the best results when classifying time slices overall?

1.4 Research Method

2

Data Preparation

Since the KM3NeT Neutrino Telescope is still under construction, simulated data provided by Nikhef was used for this project. The data itself was split onto two parts namely *events* and *noise*, both of which came from different sources and in different formats.

The *events* dataset was provided as a *HDF5* (Hierarchical Data Format) with a size of 42MB consisting of the `/data/mc_hits` and `/data/mc_info` datasets. For the purposes of this project only the `mc_hits` dataset was used. As such, it was saved as a CSV file of size 42MB for future use. An overview of the `mc_hits` dataset (here on refered to as the `hits` dataset) is provided in Table **??**.

The *noise* data was generated using a Python library written and maintained by Nikhef, `k40gen`. `k40gen.Generators(21341, 1245, [7000., 700., 70., 0.])` was used to create an instance of a generator where the first two arguments are random seeds followed by a list of rates at which single, double, triple and quadruple hits should be generated. The generator instance is then passed into `k40gen.generate_40()` method which returns a (4, n) array containing (time, dom_id, pmt_id, tot). The position coordinates (ie. `x`, `y`, `z` coordinates) for each datapoint was provided in a *positions.detx* file which was parsed using the Numpy Python package (6) and added to the *noise* array. The Python library Pandas (7) is used to convert the array into a (n, 4) dataframe and save it as a 3.9GB CSV file. Table **??** presents an overview of the `noise` dataset.

To create the final dataset for the project, *events* and *noise* need to be combined. Both datasets were read into memory as Pandas dataframes and their columns were renamed consistently. The two dataframes were concatenated and sorted based on the `time` column. The resulting dataframe was saved as a 3.7GB CSV file.

3

Data Exploration

Bibliography

[1] Silvia Adrian-Martinez, M Ageron, F Aharonian, S Aiello, A Albert, F Ameli, E Anassontzis, M Andre, G Androulakis, M Anghinolfi, et al. **Letter of intent for KM3NeT 2.0**. *Journal of Physics G: Nuclear and Particle Physics*, **43**(8):084001, 2016. 1, 2

[2] Sebastiano Aiello, Fabrizio Ameli, Annarita Margiotta, Michel Andre, Giorgos Androulakis, Marco Anghinolfi, Antonio Marinelli, Gisela Anton, Miquel Ardid, Christos Markou, et al. **KM3NeT front-end and readout electronics system: hardware, firmware, and software**. *Journal of Astronomical Telescopes, Instruments, and Systems*, **5**(4):046001, 2019. 1, 2

[3] Annarita Margiotta, KM3NeT Collaboration, et al. **The KM3NeT deep-sea neutrino telescope**. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **766**:83–87, 2014. 2

[4] Maarten Post. *"KM3NNeT" A neural network for triggering and classifying raw KM3NeT data.* PhD thesis, Universiteit van Amsterdam, 2019. 2

[5] Konrad Karaś. *Data processing pipeline for the KM3NeT neutrino telescope.* PhD thesis, Universiteit van Amsterdam, 2019. 2, 3

[6] **NumPy - The fundamental package for scientific computing with Python**. 5

[7] **Pandas**. 5