ARTIFICIAL INTELLIGENCE

The CERN Laboratory Particle Accelerator: Coping With the Analysis of Big Data

by
Mohammed Alsadah
201913930—214-42/10
Major: Physics

for
Mr. Steven Lees
Academic & Professional Communication
English 214

Abstract

This report examines how CERN laboratory is analyzing big data generated from the Large Hadron Collider by using a subfield of artificial intelligence called machine learning. In addition, it will shed light on the newly devised algorithms that can speed up the analysis.

26 October 2021

Table of Contents

INTRODUCTION 3
I. THE LARGE HADRON COLLIDER 3
A. An Examination of the Acceleration Process at the LHC 3
B. A Description of Particle Detector 4
II. BIG DATA 4
A. Definition of Big Data 4
B. The Generation of Big Data at Particles Detectors 4
III. THE MARRIAGE OF TWO FIELDS 4
A. Data Analysis With Machine Learning 5
1. Definition and approaches 5
2. Methods and algorithms 5
B. CERN Investing in Machine Learning 5
C. Acquiring New Machine Learning Methods 6
IV. RECOMMENDATIONS 6
CONCLUSION 6
REFERENCES 8

INTRODUCTION

This report will expose the reader to the challenge that the European Organization for Nuclear Research (CERN) is facing when collecting data. CERN is the world's biggest organization in the field of particle physics or sometimes called high energy physics. Its main interest is to discover new particles which enable physicists to understand and classify the elementary constituents of matter and how they interact with each other. The largest discovery of CERN laboratory is the verification of the existence of the elementary particle called the Higgs boson to which its theoretical discovery was awarded the Nobel prize. CERN laboratory has the longest particle accelerator called the large hadron collider (LHC) with a circumference of about 27 kilometers. The large hadron collider can be used to accelerate particles to near the speed of light and make them undergo head-on collisions to generate other new particles. Particle detectors at CERN use trigger systems to look at whether to discard or keep the data collected from the accelerator.

The problem with such a decision lies in making it for a large amount of data in a very limited amount of time or otherwise, particles will decay and nothing interesting will be observed. To handle and deal with this problem CERN uses a branch from artificial intelligence called machine learning which trains Computers to speed up this process. The aim of this report is to breakdown the methods used to overcome this challenge.

The talk will be restricted to CERN laboratory and its accelerator and not any other laboratories. In addition, the report will only explain the problems and solutions encountered by CERN organization and will also analyze them. Neither new solutions will be offered, nor the implementation of the suggested methods will be treated to avoid technicality. In accordance with the previous points, this report addresses the educated reader in any field.

I. THE LARGE HADRON COLLIDER

A. An Examination of the Acceleration Process at the LHC

The Large Hadron collider is located underground in a tunnel that is filled with magnets. There are two beam pipes for which each carries a beam of particles that are accelerated in opposite direction. Collisions or the intersection of beams happen at four points. The acceleration of the particles and the bending of the beam are done using electromagnets in which the electric field produces a magnetic field. The electric field provides the linear motion of the particle and the magnetic field bend that path to be circular and guides the beam through the pipe. According to Myers (2013), the magnets in the LHC have three tasks. The first is to bend the beam using dipole magnets and second is to focus the beam to increase the chance of collision and the third is to correct deviations in the previous tasks.

B. A Description of Particle Detector

There are particle detectors around the accelerator that identify particles of interest by looking at their properties such as charge, speed, or mass. Various discoveries were enabled because of particle detectors. With the help of detectors, developments in high energy physics and nuclear physics have been achieved through measuring radiation and particles properties (Kolanoski & Wermes, 2020, p.3). Different types of particle detectors are available at the LHC. The tracking device measures particle's momentum through its curved path and the calorimeter device is used to determine its energy. Collisions of particles could reach 1 billion. The data that results from collisions can reach in one day to 1PB (10^{15} Bytes). Which exposes CERN to the question of what to keep and what to leave.

II. BIG DATA

A. Definition of Big Data

The process of analyzing and manipulating a large amount of data and extracting meaningful information from a dataset with non-standard analysis techniques is called big data. Big data can offer obstacles in sharing, storing, and managing data. There are a lot of characteristics other than volume such as velocity and value that can determine whether the data is big or not since volume is a relative concept for different organizations. Velocity refers to the speed of data production and value indicates how much profitable the analysis is. Moreover, big data can take two forms, structured and unstructured. As stated by Segal (2021), the data that comes in numeric format is structured and is easy to deal with in contrast to unstructured data which comes in a qualitative format like a text or a video and requires more complex techniques for analysis.

B. The Generation of Big Data at Particles Detectors

The data generated from bombarding particles are so enormous that dealing with them with traditional methods will not suffice. In data analysis, this Is referred to with the term big data. Gleyzer et al. (2018) stated that proton collisions in the LHC can produce data of about 25GB every second and consequently, making CERN passes 200PB of stored data. To solve this problem, CERN started a project by collaborating with 42 countries to build a computing grid. A computing grid is made to achieve a common goal through the usage of computer distributed resources (like devices or files). The project is called the Worldwide LHC Computing Grid (WLCG) which comprises 170 centers in different countries. This made the processing task much easier and enabled physicists in any part of the world to share and access data.

III. THE MARRIAGE OF TWO FIELDS

A. Data Analysis With Machine Learning

1. Definition and approaches

Machine learning is defined as the process of training machines to do a certain task using data. This could be done in different approaches, two of which are supervised and unsupervised learning. In supervised learning, the device is fed a dataset with inputs and their corresponding outputs. The goal then is to discover a pattern or predict a rule that correlates an input with an output from new data based on the old one called the training data. The other approach is unsupervised learning in which the machine is only given inputs without outputs and the purpose is to collect the data in groups with the same characteristics. Products on Amazon or videos on YouTube are suggested to the user through machine learning algorithms. According to Duggal (2021), product recommendations are done by tracking the user behavior using machine learning and based on the customer's purchase history and searching patterns.

2. Methods and algorithms

For the case of physics, particle physicists are one of the early users of machine learning in analyzing data. It is used in the classification of events (collisions) which can be events of interest or not. Events that should be analyzed are called signals and the ones that should be discarded are called backgrounds. Bourilkov (2019) mentions that Due to the immense backgrounds from the LHC, extracting signals became difficult. This put the need to introduce machine learning methods for classification tasks. Such algorithms are supervised, and the outputs are classes such as backgrounds or signals. To solve such a problem, CERN created a software for data analysis called ROOT and in it, one can find a package called TMVA that implements machine learning tools that can achieve the classification task.

The other information that physicists want to obtain from the data is the track of the particles going through a detector called a tracker. This can be done by recording electrical signals from charged particles and trying to associate these signals with a single track. After that, a curve fitting process called regression (which is a common supervised machine learning algorithm) is applied to find the path of the particle. From the track of the particle, physicists obtain its momentum.

B. CERN Investing in Machine Learning

To fulfill the goal of reconstructing tracks of particles quickly, CERN organized a competition on Kaggle platform and offered the winners prizes of 25000 dollars. The competition name is TrackML which refers to the use of machine learning techniques to reconstruct the track. This competition is created to stir both the physics and computer science community to push the boundaries of current knowledge and help in the discovery of new particles. Amrouche et al. (2021) state that the competition was successful because of the simplified but realistic design of the problem that enabled computer scientists to participate.

C. Acquiring New Machine Learning Methods

CERN is working on a project called the High-Luminosity Large Hadron Collider (HL-LHC) which aims to upgrade the LHC and take it to the next level. The project is scheduled to finish in 2026 and its purpose is to increase the luminosity of the accelerator. Luminosity is defined to be the number of collisions happening at a certain amount of time. Consequently, the next upgrade will increase data gathered from the detectors if luminosity were to increase. Lopes (2021) indicates that with the enhanced LHC, the rate of collisions is going to increase by a factor between 5 and 7 and therefore physicists are looking for new machine learning algorithms that can speed up the decision making of the trigger system. A method called deep neural network is used to make the trigger system decide in a short period of time. This method resembles the nervous system of the human body. The units of the network can be thought of as neurons and the connection between them are like synapses that communicate signals. A bunch of units is collected in a layer and if a network has an enormous amount of layers, it is called deep neural network.

IV. RECOMMENDATIONS

Some lessons could be drawn from this article and should be applied by the scientific community, and these are the following:

1. Collaborative work should be emphasized as it saves time, effort, and resources and enables what could not be done individually. An example observed earlier is the Worldwide LHC Computing Grid which is distributed in 170 centers over 42 countries.

2. There should be no barriers between different fields of science as they are complementary to each other. In addition, one could find insights from one field and apply them to other fields like artificial intelligence and physics as discussed previously, or biology and chemistry or computer science with mathematics.

3. The data era has come and so companies should invest and look for methods to analyze data effectively to extract precious information like the investment CERN did to deal with big data.

CONCLUSION

To conclude, the Large Hadron Collider is a particle accelerator that accelerates particles to near the speed of light and makes them collide. The collisions detected by particle detectors are so enormous and generate large volumes of data which is referred to by the term big data. This raised CERN’s need to cope with and solve such a problem. The first step was creating the project called the Worldwide LHC Computing Grid which was a collaborative project between several countries that resulted in providing a computational power and data access from any place in the world. The second step was to use computational intelligence namely, machine learning algorithms to do the analysis. CERN used machine learning for the classification of events and track reconstruction. To find solutions to the problem of track reconstruction, CERN conducted a competition and offered participants prizes. Recently, CERN is upgrading its accelerator which will yield data more than before, and thus it needs to speed up the analysis. This is done using an algorithm called deep neural network.

REFERENCES

Amrouche, S., Basara, L., Calafiura, P., Estrade, V., Farrell, S., Ferreira, D. R., Finnie, L., Finnie, N., Germain, C., Gligorov, V. V., Golling, T., Gorbunov, S., Gray, H., Guyon, I., Hushchyn, M., Innocente, V., Kiehn, M., Moyse, E., Puget, J.-F., … Yilmaz, Y. (2019). The Tracking Machine Learning Challenge: Accuracy Phase. *The NeurIPS '18 Competition*, 231–264. https://doi.org/10.1007/978-3-030-29135-8_9

Bourilkov, D. (2019). Machine and deep learning applications in particle physics. *International Journal of Modern Physics A*, *34*(35), 1930019. https://doi.org/10.1142/s0217751x19300199

Duggal, N. (2021, October 14). *Top 10 machine learning (ML) applications & examples in 2021*. Simplilearn.com. Retrieved October 31, 2021, from https://www.simplilearn.com/tutorials/machine-learning-tutorial/machine-learning-applications.

Gleyzer, S., Carminati, F., Vallecorsa, S., & Gallix, D. (2018, July 9). *The rise of deep learning*. CERN Courier. Retrieved September 28, 2021, from https://cerncourier.com/a/the-rise-of-deep-learning/.

Lopes, A. (2021, June 21). *Speeding up machine learning for particle physics*. CERN. Retrieved September 28, 2021, from https://home.cern/news/news/physics/speeding-machine-learning-particle-physics.

Myers, S. (2013). The large hadron collider 2008–2013. *International Journal of Modern Physics A*, *28*(25), 22. https://doi.org/10.1142/s0217751x13300354

Segal, T. (2021, September 13). *Big data*. Investopedia. Retrieved October 31, 2021, from https://www.investopedia.com/terms/b/big-data.asp.

Wermes, N., & Kolanoski, H. (2020). *Particle detectors: Fundamentals and applications*. Oxford University Press.