Discovery of new subatomic particles through machine learning algorithms in CERN databases

Instituto Federal de Educação, Ciência e Tecnologia Fluminense - IFF Campos dos Goytacazes
Address: Street Dr. Siqueira, 273 - Dom Bosco Park, Campos dos Goytacazes - RJ - BRAZIL, CEP 28030-130 Tel .: +55 (22) 2726-2800

Abstract

Scientists at the European Organization for Nuclear Research, CERN, are colliding protons and re-creating true mini big bangs to find out what our universe is made of. These collisions are meticulously observed by silicon detectors. If it were not enough to be orchestrating and observing such collisions, the research center still produces enormous amounts of data from these experiments, and analyzing them is a great challenge. At each collision or event, there is a large volume of data that is already in the annual petabytes scale, needing to be filtered by the scientists. As detector resolution improves, it is necessary for increasingly robust software to be developed for real-time processing in order to filter out the most promising events, invariably producing more data. To help solve this problem, experts around the world are developing solutions based on machine learning algorithms because of their great ease in handling large volumes of data. These algorithms are familiar with problems that target, pattern recognition, regression, clustering, among others. Many of the data already captured by the hadron collider is available to the public by CERN itself. This research project aims to use these algorithms in order to study and create models left by the trajectory of these particles to find patterns that point to new discoveries (SPRACE, [s.d.]).

Objectives of the research and description of the problem

The objective of the research is the use of algorithms of machine learning in the detection of new particles. This research is based on the premise that new elementary particles can be detected by testing hypotheses not predicted by the Standard Model (ELTON ALISSON, [s.d.]). CERN continues to invest in the improvement of its detectors thus making all orders of magnitude with the surrounding increase steadily. Improvements in the magnetic fields that guide the protons, for example, have increased the number of collisions. Invariably, where it has more collisions, it has more data, so methods of classification and categorization are becoming obsolete when working with huge volumes of data that are emerging. The solution to such obsolescence is to deliver these problems to the artificial intelligence, as they are commonly called machine learning algorithms. This algorithms are excellent for handling large amounts of data, and through their trained models, are able to find patterns and rebuild models, for example. Fields of science such as medicine and astronomy already use such tools, particle physics will be just another example to follow (Devin Coldeworth, 2018).

Among the great challenges surrounding CERN, the storage of data is one of them. Companies like Oracle manage their databases often and have a set of tools for storage and availability. Each year of operation of the Large Hadron Collider (LHC), the 4 particle detectors capture trillions of collisions and accumulate about 20 Petabytes (PB) of data that are managed by Oracle tools. These technologies provide the necessary functionality for high availability, scalability, performance, distribution, protection and data management ("NetApp Search Results", [s.d.]). It is precisely this amount of data that drives a solution towards artificial intelligence and not to old projects and solutions.

Description of the work to be done and expected results

The work will be done through the development of algorithms of machine learning in databases made available by CERN in search of new trajectories and/or particles or establish some already known. Some algorithms are already commonly used by the research center, they are, Logical Regression, Decision Trees, Naive Bayes, Neighboring Neighbors (KNN), Linear Regression, Neural Networks among many others. It will be through the development of these algorithms that the research will be based.

The expected results may be from the discovery of a new particle, predicted or not by the MP, as a modest understanding of this model, thus possibly helping the synthesis of new theories.

Relation of this work with previous ones of the same area

The work developed by this research program will relate to other works of the same area when both deal with large volumes of data. Currently, solutions to problems derived from enormous amounts of data are being developed from artificial intelligence, whose machine learning belongs. Data Science, as it is widely known, is a field of science that develops in conjunction with today's great human needs, and is therefore of proven relevance when used by one of the world's largest research centers, CERN.

Data policies

Like most of the scientific computing community, I intend to make available all the knowledge developed in this research in order not only to be grateful for all the knowledge that I have already acquired for free by numerous means of communication, but also to be able to add some of the knowledge developed by me from these. In this way, others can acquire knowledge in a clear, objective and freeway, just like me. I will use both versioning software for publication of codes and availability in scientific knowledge databases of all material developed by this research.

References

SPRACE. **CERN lança desafio de Inteligência Artificial | sprace.org.br**. Disponível em: <https://www.sprace.org.br/divulgacao/noticias/cern-lanca-desafio-de-inteligencia-artificial>. Acesso em: 23 maio. 2018.

DEVIN COLDEWEY. **Particle physics gets the machine learning treatment as collider data multipliesTechCrunch**, 4 maio 2018. Disponível em: <http://social.techcrunch.com/2018/05/04/particle-physics-gets-the-machine-learning-treatment-as-collider-data-multiplies/>. Acesso em: 23 maio. 2018

ELTON ALISSON. **Cientistas planejam detectar novas partículas elementares no LHC a partir de 2015**. Disponível em: <http://agencia.fapesp.br/cientistas_planejam_detectar_novas_particulas_elementares_no_lhc_a_partir_de_2015/19661/>. Acesso em: 23 maio. 2018.

NetApp Search Results. Disponível em: <https://www.netapp.com.br/search/index.aspx?for_cr=brasil&q=CERN+%E2%80%94+o+centro+europeu+de+pesquisa+nuclear>. Acesso em: 23 maio. 2018.