Machine Learning at the ICCUB

Monday, 23 October 2023 - Tuesday, 24 October 2023

UB Physics Faculty, Sala de Graus Eduard Fontserè

Book of Abstracts

Contents

1

Quasar classification and resdhift estimation

Author: Ignasi Pérez[1]

[1] *ICCUB*

Corresponding Author: iprafols@fqa.ub.edu

I will present SQUEzE, a machine-learning-based quasar classifier that assigns redshifts to the classified objects. SQUEzE was originally designed to work with SDSS (DESI and WEAVE too) spectra but is highly flexible and can also cope with photometric data from multi-narrow-band photometric surveys (e.g. J-PAS). It follows the human visual inspection procedure and classifies the objects using random forest classifiers based on the found emission line peaks. The redshifts are automatically extracted from the position of the peaks.

2

Neural Network classifier for the generation of clean Magellanic Cloud samples

Author: Óscar Jiménez Arranz[None]

Co-authors: Eduard Masana [1]; Laurent Chemin [2]; Merce Romero-Gomez ; Paul McMillan [3]; Teresa Antoja [4]; Xavier Luri [5]; santi Roca-Fàbrega [6]

[1] *Institut de Ciències del Cosmos (ICCUB-IEEC)*

[2] *Instituto de Astrofísica*

[3] *Lund Observatory*

[4] *Institut de Ciències del Cosmos*

[5] *ICCUB*

[6] *Universidad Complutense de Madrid*

Corresponding Authors: paul@astro.lu.se, xluri@fqa.ub.edu, mromero@fqa.ub.edu, astro.chemin@gmail.com, sroca01@ucm.es, emasana@fqa.ub.edu, tantoja@fqa.ub.edu, ojimenez@icc.ub.edu

The proximity of the Magellanic Clouds (MCs) to the Milky Way (MW) makes them a perfect laboratory for testing methodologies and models designed for the study of external galaxies using Gaia (ESA) data. To do so, we need to separate in the Gaia data the MCs stars from the foreground MW stars, in order to obtain "clean"MC samples. This is achieved through the design and training of a neural network classifier; the algorithm uses as much of the Gaia DR3 types of data as possible and its results have been validated through comparison with independently classified data (Jiménez-Arranz+23a,b).

3

Search for lensed gravitational-wave events with machine learning

Authors: Roberto Bada Nerín[1]; Oleg Bulashenko[2]

[1] *UPC, Barcelona*

[2] *UB*

Corresponding Authors: rbadan99@gmail.com, oleg@fqa.ub.edu

The growing significance of Gravitational Wave Astrophysics puts in evidence the need of techniques capable of effectively and reliably analyzing all the collected data. Furthermore, the search for lensing signatures within gravitational-wave signals is a challenging task that holds the potential to uncover fresh insights into fundamental physics, astrophysics, and cosmology. In this context, we train a set of different neural networks with various types of Gravitational Wave data and we compare their performance when classifying the data in three groups: plain noise, unlensed signal and lensed signal

4

Real vs. bogus classifier in time-domain surveys

Author: Nadia Blagorodnova[1]

Co-author: Fiorenzo Stoppa [2]

[1] *Univeristat de Barcelona*

[2] *Radboud University*

Corresponding Authors: nblago@fqa.ub.edu, f.stoppa@astro.ru.nl

Time-domain surveys are designed to study astrophysical transient phenomena appearing in the night sky. The improvements in instrumentation and data analysis are allowing the new generation of surveys to discover several thousand (and soon to be millions) of events per night. However, some of such discoveries are associated with spurious detections related to spikes from bright stars, parasitic optical reflections, cosmic rays, or defects in the detectors, just to name a few. These spurious or "bogus" detections complicate the arduous task of finding interesting real events in such a big "stream" of events. To help human scientists, several time-domain surveys have implemented Machine Learning approaches designed to automatically classify the detections as "real" or "bogus". I have contributed to the creation of real-bogus classifiers for two different time-domain surveys. In my talk, I will describe the challenge, the main algorithms used by each survey, and the main lessons learned from our efforts to compile a meaningful training set.

5

Machine learning in the context of large stellar surveys

Author: Friedrich Anders[1]

[1] *Universitat de Barcelona*

Corresponding Author: fanders@icc.ub.edu

The Gaia mission as well as large-scale ground-based spectroscopic surveys are collecting complex data for millions (even billions) of stars. Within the Gaia group we are therefore been using more and more machine-learning methods to cope with the amount of data. In this talk I will present some examples from recent publications in which we have successfully used supervised (typically regression) and unsupervised (typically dimensionality reduction and clustering) methods in the context of stellar and Galactic astrophysics.

6

Machine learning in LHCb

Author: Lukas Calefice[1]

[1] *Universitat de Barcelona/ICCUB*

Corresponding Author: lukas.calefice@cern.ch

Machine learning techniques have a variety of use cases within the LHCb experiment. They are an essential ingredient to achieve the ultimate performance in event reconstruction and high precision in physics output. This talk will give an insight to the use of ML algorithms in online event selections performed by the LHCb trigger system, offline data analyses of physics measurements, as well as to track and electron reconstruction and particle identification algorithms. Finally, the ML use in the LHCb simulation framework and the experience of the UB LHCb group with machine learning in LHCb will be discussed.

7

Symmetries in Neural Quantum States

Author: Javi Rozalén Sarmiento[1]

[1] *Universitat de Barcelona*

Corresponding Author: jrozalen@ub.edu

Neural Quantum States are at the basis of a new ab-initio method especially designed to tackle the quantum many-body problem. These combine the variational method with neural networks, a flagship tool of modern Machine Learning. Neural Quantum States have been successfully used in spin, electronic and nuclear many-body systems. Neural networks can provide an unbiased approximation of complex wave functions, and so far the growth of network parameters with particle number has been found to be polynomial. One expects the growth to be further mitigated by restricting ansätze to the manifold of states that respect physical symmetries. To this end, starting from the most general way to make a neural network equivariant to a certain symmetry group, we design a many-body neural network ansatz which respects the fermionic particle-exchange symmetry. Previous fermionic ansätze are a specific case of this general approach, which we develop formally, foreseeing the need to develop relevant nuclear symmetries, like spin and isospin, into a Neural Quantum State. I will discuss how this approach will be exploited in future nuclear physics simulations and show some initial results on test systems.

8

Anomaly detection for non-gaussian transient noise characterization in LIGO data using auxiliary channel information

Author: Paloma Laguarta[None]

Corresponding Author: plgonzalez99@gmail.com

Gravitational-wave interferometers are able to detect a change in distance of ~ 1/10,000th the size of a proton. Such sensitivity leads to large rates of appearance of non-gaussian transient noise bursts in the main output of the detectors (the strain), also known as glitches. These glitches come in a wide range of frequency- amplitude-time morphologies and have unknown environmental and instrumental causes, hindering searches for gravitational wave transients. Current Machine Learning approaches for studying glitches use their strain morphologies to classify them but do not consider relevant information provided by auxiliary channels that are distributed through the detectors and monitor their state. In this proof-of-concept work, an unsupervised approach is taken by using the auxiliary channel data encoded in a fractal dimension measure to train an autoencoder with circular

convolutions in an anomaly detection fashion. Basing the analysis on the autoencoder's embedded space, this methodology uncovers unknown glitch morphologies, overlaps between different glitches, and misclassifications by the current state of the art.

9

A four philosophers' introduction to machine learning (Invited talk)

Author: Oriol Pujol[1]

[1] *U Barcelona (Math & CS)*

AI is everywhere and suddenly everybody is an expert on machine learning. In this talk, I will try to give a gentle introduction the main topics in the design of machine learning algorithms and what guarantees we have they will work. The relationship between ML and empirical science method will be highlighted and some applications to the Physics domain discussed.

10

Machine Learning Techniques in Strong Gravity (Invited talk)

In this talk we will review some recent uses of machine learning techniques to perform calculations in strong gravity. These will include physics-informed neural networks (PINNs) for the solution of differential equations, and generative models such as GANs.

12

Transforming Healthcare: Machine Learning and Computer Vision at the Forefront (Invited Talk)

Corresponding Author: gemma.piella@upf.edu

13

Data Science Approaches to Physics and Complex Systems (Invited talk)

As physicists, we know that experiments can produce massive amounts of data. However, nowadays, we collect information from many different sources. We store different kinds of data points when we run simulations. Experiments do save data beyond what they are designed to provide (for example, metadata or instrument data). And, in complex system science, data comes from everywhere, from social networks to climatological data. Data science approaches can help generate new opportunities and exploit the vast amount of data the research produces. In this seminar, I will present some examples that go from how machine learning explanations can be used to better understand simulations to an augmented reality app that allows us to experiment and explore the human role in segregation.

14

Registration