A novel machine learning approach to delve into the AGN central engine

MiniGrant RSN1

Participants: Susanna Bisogni, A. Gargiulo, M. Scodeggio (INAF-IASF Milano)

External collaborator: Dr. Umberto Michelucci

Scheda: LSST@IASF-Milano “Galaxy and AGN Evolution Studies with Large Spectroscopic Surveys” (PI: M. Scodeggio)

Budget: 15.000 €, costs for visits with external collaborator/conferences

Purpose: The scientific goal of this project is the study of the physics of accretion in AGN and of its connection with the powerful nuclear outflows, responsible for triggering the energy transfer from the sub-pc to the kpc scales. This is addressed through the analysis of the X-ray (hot-corona) and UV/optical (accretion disc and nebular regions, e.g. BLR and NLR) emissions and their correlations. The financial request of this project is aimed at covering costs for travelling of proponents and collaborator for reciprocal visits and costs for conferences. The final goal is defining a tailored strategy to extract, through appropriate machine learning techniques, all the information stored in the unique data set we collected and to isolate main trends between the spectral properties of disc, corona and nebular regions. This will allow us to set constraints on the physics of accretion and nuclear winds.

First phase: analysis of spectral properties in X-ray and UV/optical bands

Deliverable	Milestone
Project starts	T0
Table with spectral properties measurements. Spectral properties to be ingested by the machine learning algorithms for inference of the variance drivers, e.g fluxes, luminosities, slopes of continuum, line profile shape parameters, asymmetry coefficients for line profiles, velocity shifts w.r.t. the nominal wavelength, line widths, equivalent widths etc. The proposer is experienced in spectral fitting of UV/optical quasar spectra and in the measurements described above (Bisogni et al. 2017a, 2017b).	T0+6m

In the first six months of the project, I collected a unique sample of **5130 quasars** with **spectroscopic data both in the UV/optical band (SDSS DR16)** and in the **X-ray band (Chandra Source Catalog 2.0)**.

I performed a fit to the **X-ray spectra**, to estimate the flux and slope of the continuum emitted by the hot corona in the AGN. To ensure consistency in the analysis of a sample with widely varying S/N, I applied a model consisting of a simple power law corrected for Galactic absorption to all the sources. This model allows us to retrieve **fluxes** and **photon index** that provide information on the amount of energy produced in the central engine and on the slope of the SED, which is crucial to studying the ionisation in the surrounding nebular regions.

For the analysis on the **UV/optical spectra**, we started from the fitting models by Wu & Shen 2022 and measured quantities for both continuum and emission lines. Slopes and fluxes again provide info on the engine's output, while quantities measured on the lines (e.g. flux, velocity shifts, widths, asymmetry indexes) are used to study the kinematics of the nebular regions and the presence of winds, which may be responsible for the energy transfer to the galactic scales.

Second phase: machine learning analysis

As to November 2023, we are on schedule with the planning presented in the request

Deliverable	Milestone
Routine in Python/Jupyter notebook (machine learning analysis). Requires the interaction between proposer, expert and collaborators though visits	T0+12m

In the second, ongoing part of the project, I'm focusing on two aspects:

- the collaboration with Dr. Michelucci (currently online only) to identify, define and design the best machine learning algorithm that is suited for the analysis, taking into account the peculiarities/criticalities of the sample (e.g. spectra do not cover the same wavelength range and not all bands/lines are available for all the sources. This is somehow related to the problem of missing values in machine learning and needs to be properly addressed algorithmically). The first in person visit of Dr. Michelucci at INAF-IASF Milano is planned for the end of November 2023 (~1 week)
- the organisation of a cycle of lessons of machine and deep learning open to INAF researchers to be held at INAF-IASF Milano. The school is scheduled for the week of January 15-19, 2024. Details on the program can be found at the webpage https://www.iasf-milano.inaf.it/machinedeeplearning/

Expenses incurred so far in 2023:

- Participation to the HEAD 20 / AAS (26-30 March 2023): ~400 €
- First visit of Dr. Michelucci (November 2023): ~1000 €
- Cycle of lessons Machine and Deep Learning @IASF-Milano (January 2024): ~2000 €

At the moment, I do not report any critical issues