Abstract

The exploration of Active Galactic Nuclei (AGNs), the highly energetic centers of galaxies hosting supermassive black holes, has been a cornerstone of modern astrophysics. Milestones such as the discovery of quasars in the 1960s and the identification of the first AGN in a Seyfert galaxy have established AGNs as key tools for studying galaxy evolution and black hole growth. Despite decades of research, the mechanisms linking AGN properties to cosmic distances remain untouched. This project investigates the evolution of Type 2 AGNs across the redshift spectrum, where redshift serves as a critical measure of distance and cosmic history. We utilized data from NASA's HEASARC, specifically the XMM-CTY2AGN database, comprising multi-wavelength spectroscopic observations of 255 Type 2 AGNs. Employing advanced machine learning models like Gaussian processes and x-y kernel methods, we analyzed non-linear relationships between redshift and AGN properties. Our findings reveal that Type 2 AGNs at higher redshifts exhibit significantly greater luminosities and elevated U-V color indices. These results suggest an evolutionary trend in AGN characteristics, shedding light on their role in shaping the early universe and advancing our understanding of supermassive black hole activity.

Rationale

The study of Type-2 Active Galactic Nuclei (AGNs) is crucial in the field of astronomy to understand how supermassive black holes (SMBHs) and their host galaxies evolve throughout cosmic history. Unlike Type-1 AGNs, Type-2 AGNs are obscured by a dense, dusty torus that makes their properties more challenging to observe for us. However, they still play an equally significant role in constructing a comprehensive view of AGN populations. Our research will examine how redshift correlates with key properties of Type-2 AGNs—such as X-ray luminosity, bolometric luminosity, stellar mass, and U-V color indices—to identify trends that shed light on their role in galaxy evolution. Redshift, a central concept for understanding this study, refers to the stretching of light to longer wavelengths as it traverses the expanding universe. Higher redshift values correspond to greater distances and earlier times in the universe, enabling us to study galaxies and AGNs as they appeared billions of years ago. By analyzing the relationship between AGN properties and redshift, we want to understand the accretion processes, star formation histories, and structural transformations of these systems. This work will be particularly significant for the field because it addresses a key limitation in existing astrophysical research, which often neglects obscured AGNs, leading to an incomplete understanding of the universe's structure and evolution. Ultimately, our research has the potential to significantly improve our understanding of how supermassive black holes and their host galaxies co-evolve over time. This study can help us develop more accurate models of galaxy evolution, and by deepening our understanding of more obscured AGNs, we can improve estimates of total energy output from AGNs across the universe. With our improved understanding of high-energy astrophysics in general, our findings can also be used in the direction of improving telescopes and similar technology in the future.

Research Question

How do key properties of Type-2 AGNs correlate with redshift to help us better understand their properties and evolution?

Hypotheses

We hypothesize that galaxies with higher redshifts will generally exhibit higher stellar masses, greater bolometric and X-ray luminosities, and larger U-V magnitudes. This is because we understand that galaxies in the early universe were more actively accreting material, resulting in higher luminosities. Additionally, redder U-V magnitudes at high redshift indicate older stellar populations and reduced star formation activity, while higher stellar masses are expected as galaxies accumulate more stars over time. However, it's important to note that the evolution of these properties is more complex than it would initially seem, as it is influenced by various external factors such as accretion rates, galaxy interactions, and cosmic environment.

Engineering Goal

We will utilize machine learning models to analyze the relationship between redshift and the physical properties of Type-2 AGNs to better understand their evolution and properties, particularly accounting for non-linear relationships. This approach will allow us to identify key trends and also account for uncertainties in the data. This will be our method for studying AGN evolution over cosmic time.

Materials and Procedures

To begin the analysis, we will source our data from NASA's High Energy Astrophysics Science Archive Research Center (HEASARC), specifically the XMM-COSMOS X-Ray Selected Type-2 AGN catalog. This dataset contains information on 255 spectroscopically identified AGNs, focusing on the variables such as X-ray luminosity, bolometric luminosity, stellar mass, and U-V magnitudes. We will perform our analysis using Python in Jupyter Notebook, using the Visual Studio Code environment. Key libraries will include Pandas and Numpy for data manipulation, Scikit-learn for machine learning, and Matplotlib and Seaborn for data visualization.

To clean the dataset, we will first remove unnecessary columns using the drop function and address missing values by replacing NaN entries with column means computed using the .mean function. This preprocessing step will ensure a consistent and reliable dataset for analysis. After that, we will split the data into features (independent variables) and the target variable (redshift). We will further partition the dataset into randomized training (80%) and testing (20%) sets.

We'll also be employing the Gaussian Process Regression (GPR) with a kernel combining DotProduct and WhiteKernel. The DotProduct kernel is optimal for linear modeling, while the WhiteKernel accounts for noise in the data. After training the model on the training data, we will evaluate its accuracy using mean squared error (MSE) and R^2 values. Finally, after all that, we'll be able to visualize the results using partial dependence plots to illustrate the influence of individual features on redshift, and linear regression to identify feature importance.

Risk Assessment

Several risks will be considered during the research. Data quality will pose a significant challenge, with potential biases arising from the limited observations of Type-2 AGNs and errors in data extraction. To mitigate these risks, we will validate data preprocessing steps and cross-reference our findings with existing papers. Computational risks, including overfitting and the high computational demand of GPR models, will be addressed through hyperparameter tuning and efficient data splitting. It’s important to note that there will be a risk of misinterpreting correlations as causations, but we will try to mitigate this by focusing on statistically significant trends supported by proven astrophysical theories.

Data Analysis

The data analysis will employ multiple statistical models to explore the relationship between AGN properties and redshift. We plan to test several other models to employ in our research such as Lasso Regression, Random Forest Classifiers, LightGBM, and Gaussian Regression Model. From our preliminary analysis, we noted the benefits of the Gaussian Regression model such as its uncertainty quantification, its usage of Kernals to adapt the model to different types of data, and its high accuracy with smaller data sets. With this in mind, we hope to leverage the advantages of the model to create dependency graphs and feature importance graphs to determine what properties of type 2 AGNs have the greatest correlation with redshift. This could help us better understand the physical properties and evolution of AGNs across spacetime.

Bibliography

Mountrichas, G. (n.d.). AGN Host Galaxy Properties. Astronomy & Astrophysics Journal. Retrieved from https://arxiv.org/html/2009.10862v5#Sx2

NASA HEASARC Archive. (n.d.). XMM-COSMOS X-Ray Selected Type-2 AGN Catalog. Retrieved from https://heasarc.gsfc.nasa.gov/W3Browse/galaxy-catalog/xmmcty2agn.html

Pryzant, R. (n.d.). Gaussian Process Regression in Machine Learning. Stanford Blog. Retrieved from https://cs.stanford.edu/~rpryzant/blog/gp/gp.html

Schmidt, M. (n.d.). Quasars and AGN Luminosity Functions. arXiv preprint. Retrieved from https://arxiv.org/html/2009.10862v5#Sx2

Scikit-learn Documentation. (n.d.). Gaussian Process Regression. Retrieved from https://scikit-learn.org/stable/modules/gaussian_process.html

Scikit-learn Documentation. (n.d.). Gaussian Process Kernels: DotProduct Kernel. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.kernels.DotProduct.html

Scikit-learn Documentation. (n.d.). Gaussian Process Kernels: WhiteKernel. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.kernels.WhiteKernel.html
Scikit-learn Documentation. (n.d.). Gaussian Process Regression: Kernels and Applications. Retrieved from https://scikit-learn.org/stable/modules/gaussian_process.html
Additional Resources. (n.d.). AGN Evolution and Machine Learning Approaches. Retrieved from arXiv.org
Bolometric Luminosities and Eddington Ratios of X-ray Selected Active Galactic Nuclei in the XMM-COSMOS Survey
Lusso, E., Comastri, A., Simmons, B. D., Mignoli, M., Zamorani, G., Vignali, C., ... Riguccini, L. (2012). Bolometric luminosities and Eddington ratios of X-ray selected active galactic nuclei in the XMM-COSMOS survey. Astronomy & Astrophysics, 548, A106. DOI: 10.1051/0004-6361/201220258
Bolometric Luminosities and Eddington Ratios of X-ray Selected Active Galactic Nuclei in the XMM-COSMOS Survey
Lusso, E., Comastri, A., Simmons, B. D., Mignoli, M., Zamorani, G., Vignali, C., ... Riguccini, L. (2012). Bolometric luminosities and Eddington ratios of X-ray selected active galactic nuclei in the XMM-COSMOS survey. Astronomy & Astrophysics, 548, A106. Retrieved from https://ar5iv.labs.arxiv.org/html/1206.2642v1
Probing Star Formation Rates and Histories in AGN and Non-AGN Galaxies Across Diverse Cosmic Environments
Mountrichas, G., Siudek, M., & Cucciati, O. (2023). Probing star formation rates and histories in AGN and non-AGN galaxies across diverse cosmic environments and extensive X-ray luminosity ranges. arXiv Preprint. Retrieved from https://ar5iv.labs.arxiv.org/html/2404.02959v1