RESEARCH ARTICLE | OCTOBER 01 2024

The Application of Machine Learning to AGN Classification

Vivek Abraham; Joel Deville; Garv Kinariwala

J. Undergrad. Res. Phys. Astron. 34, 100001 (2024)
https://doi.org/10.1063/10.0034182

The Application of Machine Learning to AGN Classification

Vivek Abraham,[a] Joel Deville,[b] and Garv Kinariwala[c]

University of Texas at Austin
2515 Speedway, Austin, Texas 78712, USA

[a] vivek.abraham@utexas.edu
[b] jdeville@utexas.edu
[c] garvkinariwala@gmail.com

Abstract. A relatively new development in quasar and Seyfert research is the utilization of machine learning to expedite data collection and aid in analysis. This paper discusses a specific application of machine learning: to classify active galactic nuclei (AGN) as Seyfert type 1s, Seyfert type 2s, or quasars. Initially we focus on summarizing the development of research on the nuclei types from their discovery to present day. Then, our paper moves to a more focused discussion of the utilization of machine learning to classify AGN types. The importance of expedited AGN classification, as well as avenues for future research into the intersection of classification algorithms and AGNs, are discussed.

INTRODUCTION

Since their discovery, two types of active galactic nuclei (AGN) have gained significant traction in the scientific community: Seyferts and quasars. Seyferts, first identified in the 1940s, are distinguished by the broad emission lines in their spectra and emit very high amounts of radiation, usually rivaling the amount emitted by the entirety of the constituent stars in their host galaxies [1]. As observational technology progressed and research into AGNs developed, astronomers discovered that Seyferts could be further categorized into two major subclasses: Seyfert 1 galaxies, which exhibit both narrow and broad emission lines in their spectra, and Seyfert 2 galaxies, which only exhibit (relatively) narrow emission lines [2].

Quasars, first identified in the 1960s, appear as extremely luminous galactic objects. Like Seyferts, quasars emit tremendous amounts of radiation [3]. However, while Seyferts are comparable in brightness to their host galaxy, quasars greatly eclipse their hosts, outshining the rest of the galaxy by two orders of magnitude or more. Due to the parallels between Seyferts and quasars, the two AGN types are often studied in tandem. Due to their unique properties, Seyferts and quasars have been instrumental in broadening our understanding of galaxy morphology and the large-scale structure of the universe. Accelerating the rate at which we can classify these galaxies will help propel collective research efforts toward more pressing research questions regarding galaxy evolution and star formation in the early universe.

HISTORY

Research on Seyfert galaxies was spearheaded by Carl Seyfert with his 1943 paper "Nuclear Emission in Spiral Nebulae." Seyfert analyzed the spectra of six spiral galaxies that had extremely large amounts of radiation emitting from their nuclei, finding that the nuclei's spectra contained several broad emission lines that were not present in non-emitting nuclei [4]. While neither Seyfert nor the rest of the scientific community understood the mechanisms behind the nuclei type at that time, his study helped pave the way for deeper research into the phenomenon, leading to "Seyfert galaxies" being named in his honor.

By the 1960s, astronomers had discovered that Seyfert nuclei were extremely dense and had high mass, usually weighing between 10^9 and 10^{10} solar masses. In addition, the duration of peak emissions was estimated to be on the order of 10^8 years [5]. Due to increasing interest in Seyferts, a conference was held in 1968 to collate findings on the celestial objects, focusing on 13 galaxies with emission lines indicative of Seyfert activity [6]. Several theories for the broad Balmer lines found in Seyfert spectra were presented, the most popular of which postulated that the lines were produced by gas with both high electron density and high electron scattering; however, most scientists believed that these theories failed to fully explain other characteristics of Seyfert spectra, such as their very weak Bowen OIII lines [6].

As research into Seyferts steadily developed, quasars began to emerge as another enigmatic celestial phenomenon. Quasar objects were first detected in the late 1950s through all-sky radio surveys, appearing as strong, stellar-like radio emissions [7]. In 1962, the location of one such source, 3C 273, was pinpointed and its spectra analyzed. The spectra contained broad, unfamiliar emission lines and appeared to have a redshift of $z = 0.16$ [8]. While such sources were initially attributed to stars, their calculated redshifts were far greater than that of any star previously observed; moreover, interferometry determined that the angular size of the sources was very small. A year later, Maarten Schmidt hypothesized that the sources were formed through distant, extremely powerful objects [9]. However, due to the uncharacteristically massive amount of energy that the objects would have to produce to be at the distances he postulated, his theories were not widely accepted by the scientific community at the time.

While research into quasars progressed into the mid-1960s, the processes behind the phenomenon were still not well understood [9]. In 1964, Edwin Salpeter and Yakov Zeldovich suggested that quasar activity could be caused by matter falling into a supermassive black hole; however, their theory was largely dismissed because the existence of black holes was not generally accepted at that time [7]. Meanwhile, many more Seyfert galaxies were being identified; by the 1970s, more than 700 potential Seyferts were flagged [10]. As the sample of Seyfert galaxies grew, it became apparent to astronomers that Seyferts contained distinct enough spectra to warrant further classification (see Fig. 1). Daniel Weedman spearheaded the subdivision of Seyfert galaxies into Seyfert 1s and Seyfert 2s; Seyfert 1s contain broad Balmer lines (relative to the length of the forbidden lines), while Seyfert 2s contain Balmer lines and forbidden lines of approximately the same width [2].

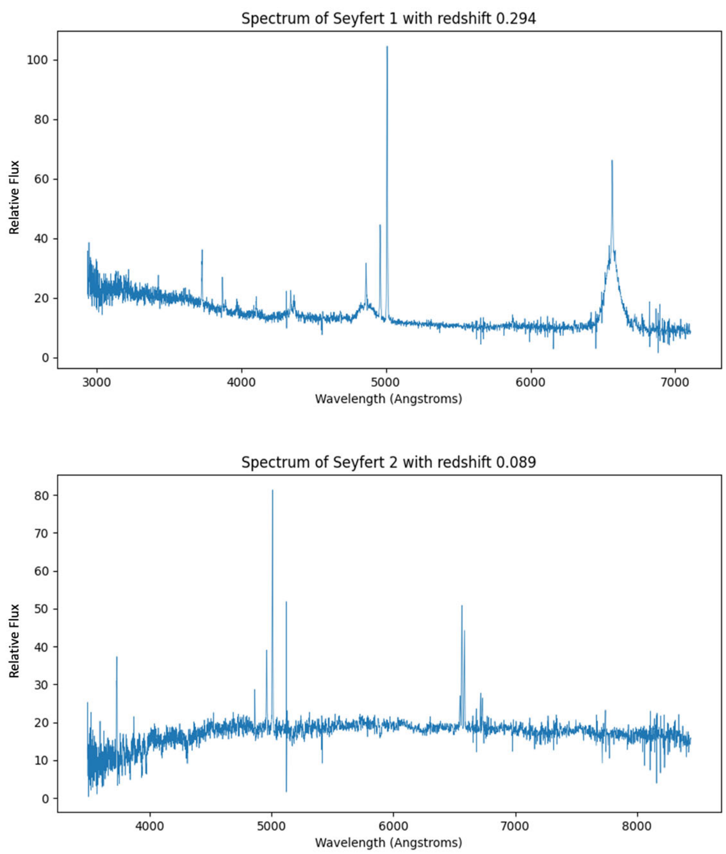

FIGURE 1. Top: Spectrum of Sloan Digital Sky Survey (SDSS) object 731880605133858816, a Seyfert 1. Bottom: Spectrum of SDSS object 435764525682157568, a Seyfert 2. Note the significantly broader Balmer lines of the top spectrum.

The existence of supermassive black holes was confirmed in 1974 by Bruce Balick and Robert Brown using a radio interferometer. The astronomers discovered evidence of a dense and immobile source emitting synchrotron radiation at the center of the Milky Way, highly indicative of a black hole [11]. The accretion disks surrounding supermassive black holes were quickly seen as key players in the phenomena observed in quasars and Seyfert galaxies. The material drawn in by the black hole spins inward, forming an accretion disk. In doing this, large amounts of energy (radiation and particles) are released [12]. By the mid-1970s, most astronomers agreed that the accretion disks surrounding supermassive black holes sufficiently explained the broad Balmer lines and extreme luminosities of quasars and Seyferts. In addition, calculations on the luminosity of matter accretion into supermassive black holes proved to be sufficient to explain the high redshifts of quasar objects [13].

More properties of Seyfert galaxies and quasars were discovered in the late 1970s. In 1977, 60 Seyfert objects were thoroughly surveyed by large-scale image-tube plates to better understand the properties of their host galaxies. While image blur and faintness made some galaxies difficult to classify, the survey appeared to show that the vast majority of Seyfert galaxies were spiral or barred spiral; very few Seyfert nuclei appeared in elliptical galaxies [14]. Meanwhile, observations of quasars revealed that most—about 90%—did not have strong radio emissions as originally believed [15]. Quasars with strong radio emissions were entitled “radio loud,” while quasars with weaker radio emissions were entitled “radio quiet.” The late 1970s also saw the discovery of the first Seyfert objects that did not cleanly fall into either the Seyfert 1 or Seyfert 2 category, warranting the need for additional Seyfert subclasses: Seyfert 1.2, 1.5, 1.8, and 1.9 [16]. As Seyfert numbering increases from 1.2 to 1.9, the broad Balmer beta line becomes increasingly weak, being nearly undetectable in Seyfert 1.9s, similar to Seyfert 2s. While the additional Seyfert subclasses warrant the need for a more careful analysis of AGN classification, most Seyfert galaxies are of type 1 or 2, and most objects in the other Seyfert subclasses can be interpolated into one of the two major classes.

The early 1980s yielded a conceptual leap in understanding Seyfert galaxies and quasars through the formulation of the unified model of AGN. Studied by Robert Antonucci, this model captured the observed diversity in AGN classes [17]. Antonucci's groundbreaking research extended into the 1990s and built upon foundational ideas proposed by Martin Rees. Antonucci emphasized the significance of the observer's line of sight concerning the toroidal gas and dust region surrounding a supermassive black hole. Seyfert 1 and Seyfert 2 galaxies share a common structure, with the differences in their observed characteristics resulting from variations in the observer's viewing angle. Radio, x-ray, and gamma-ray observations unveiled intricate AGN morphologies and energetic processes. X-ray observations, for instance, revealed the intense emissions from the accretion disks around supermassive black holes. These observations not only corroborated existing theories but also unveiled complex interactions between AGN components and their host galaxies.

The improved technology of the 1980s and 1990s yielded many advancements in quasar and Seyfert research. In addition to radio, gamma, and x-ray technology, infrared (IR) surveys from the Infrared Astronomical Satellite provided a better picture of the toroidal dust structures surrounding supermassive black holes [18]. Moreover, infrared radiation penetrates dust and gas, allowing previously obscured AGN to be detected through IR imaging. Meanwhile, ultraviolet (UV) imaging allowed for finer-grain analysis of emission-line spectra from accretion disks, as much of the continuum spectra from matter falling into supermassive black holes shows up in the UV spectrum. IR and UV imaging also revealed that Seyfert galaxies are significantly more common than previously expected, with around 16% of galaxies containing Seyfert nuclei [19]. In addition, spectroscopy and imaging allowed astronomers to identify quasar objects at extremely high redshifts. A survey of nine quasars with redshifts of $z \geq 4$ was conducted in 1995; the study found that the quasars have similar x-ray and broadband spectra to closer quasars, suggesting that the accretion dynamics of quasars have not significantly changed from a few billion years ago [20]. Due to their relative lack of change over time, quasars have been identified as being paramount to mapping the early universe and serving as long-distance standard candles [21].

In the 21st century, modern technology and more sophisticated observational techniques have allowed research on Seyferts and quasars to thrive. The Hubble Space Telescope and the Chandra X-ray Observatory have created high-resolution, multi-wavelength observations of AGNs, revealing their complicated structures and characteristics [22]. With the emergence of photometric surveys in astronomy, the challenge of processing and understanding massive image data has become increasingly apparent in the 21st century [23]. Recent research has demonstrated the power of machine learning in data processing and analysis. In the scope of AGNs, a convolutional neural network trained on quasar spectra demonstrated 99.5% accuracy in identifying quasars and estimating their redshift [24]. As technology, algorithms, and research methods continue to advance, Seyfert and quasar research remain at the forefront of astrophysics, uncovering new layers of complexity in the universe.

MACHINE LEARNING IN AGN CLASSIFICATION

One particularly interesting avenue for modern research into Seyferts and quasars involves the augmentation of machine learning as a tool to expedite classifying the AGN types. Flagging sources as potential AGN and then classifying them as quasars, Seyfert type 1s, or Seyfert type 2s are important steps in building large-scale models of AGN. Machine learning can greatly increase the speed and accuracy at which classifications can be made relative to human classification [25].

Machine learning classification can be a critical tool for identifying and categorizing celestial objects. Telescopes and observatories are constantly capturing vast amounts of data from the sky, making it increasingly challenging for astronomers to manually analyze and classify these objects. The current amount of data is staggering for humans to keep up with. For instance, the Legacy Survey of Space and Time (LSST) will produce millions of data points per night, making human-only classification infeasible [27]. On the other hand, a reasonably sized machine learning model could easily keep up with the throughput of data, given sufficient computing power. By taking advantage of machine learning, astronomers can automate the process of identifying and distinguishing between AGN types based on their spectral signatures or distinct features in images. Machine learning models may also be able to decipher trends in the raw feature sets of data that are invisible to humans. For example, a k-nearest neighbors (KNN) algorithm, given n features, can cluster data in an n-dimensional hyperplane; meanwhile, humans struggle to find patterns within data greater than three dimensions [26]. As a result, machine learning classification can help to offload and expedite the time-consuming task of data analysis.

To advance this field further, the primary obstacles to overcome relate to the number of quasars and Seyfert galaxies that we have data for. The *Veron Catalog of Quasars & AGN* has cataloged 133,326 quasars, but only 16,517 Seyfert 1s and less than 10,000 Seyfert 2s [28]. A machine learning algorithm that aims to classify quasars and Seyferts from all AGN would be hurt by the disproportionately smaller number of Seyfert objects; to combat the data imbalance, gathering more data on Seyfert objects would be beneficial. Specifically, a machine learning algorithm would be most effective when trained on a dataset that accurately reflects the actual proportions of Seyfert and quasar objects in the universe.

In addition to data, the future of this field is defined by the efficiency and accuracy of machine learning algorithms. In the future, there will likely be modifications to current popular machine learning algorithms or brand-new ones that are more efficient for this use case. Therefore, it is important to survey machine learning algorithms and hyperparameters to see if one provides better accuracy than the others. Furthering our understanding of quasars and Seyfert objects would also be beneficial for improving the accuracy of AGN classifications. By identifying features more pertinent to these objects, an algorithm would likely make more accurate inferences. One avenue for gathering more detailed data on quasars and Seyfert objects is instrumentation upgrades, which would allow us to detect more precise details on the AGN types. Using this data within any machine learning algorithm would provide the algorithm with finer granularity in determining whether or not an AGN is a quasar or a Seyfert object.

CONCLUSION

Seyferts and quasars are two highly intriguing types of cosmological phenomena, due to both the broad emission lines in their spectra and the immense radiative output they produce. While our knowledge of them has drastically improved since their discoveries in the 1940s and 1960s, respectively, our understanding of the AGN types remains incomplete. Creating and refining machine learning models to aid with the classification of broad-line spectra will help astronomers increase the efficiency of data collection on the AGN types, in turn helping us better understand the morphological evolution of galactic nuclei and the connections or differences between both Seyfert types and quasars. To make a robust machine learning model, the data imbalance between quasar objects and Seyfert galaxies needs to be remediated; in addition, more features that distinguish between AGN types need to be pinpointed. Seyferts and quasars are incredibly enigmatic phenomena that will likely continue to be at the forefront of astronomical research for many decades to come. Through careful analysis of their radiation and spectra, along with machine learning as an additional tool in our belt, we can analyze their unique properties to better understand the past, present, and future of the universe.

ACKNOWLEDGMENTS

We would like to thank the Geometry of Space stream in the Freshman Research Initiative at the University of Texas at Austin for the opportunity to pursue research as undergraduate students.

REFERENCES

1. B. Peterson, Publ. Astron. Soc. Pac. **114**, 1345–1355 (1997).
2. E. Y. Khachikian and D. W. Weedman, Astrophys. J. **192**, 581–589 (1974).
3. B. Ryden, *Quasars* (2003). Available at https://www.astronomy.ohio-state.edu/ryden.1/ast162_8/notes36.html.
4. C. K. Seyfert, Astrophys. J. **97**, 28–40 (1943).
5. L. Woltjer, Astrophys. J. **130**, 38–44 (1959).
6. A. A. Society, Astron. J. **73**, 836–841 (1968).
7. G. A. Shields, Publ. Astron. Soc. Pac. **111**, 1–37 (1999).
8. C. Hazard, M. B. Mackey, and A. J. Shimmins, Nature **197**, 1037–1039 (1963).
9. M. Schmidt, Nature **197**, 1040 (1963).
10. D. W. Weedman, Annu. Rev. Astron. Astrophys. **15**, 69–95 (1977).
11. B. Balick and R. L. Brown, Astrophys. J. **194**, 265–270 (1974).
12. B. Ryden, Accretion disks for beginners [Lecture notes] (2011). Available at https://www.astronomy.ohio-state.edu/ryden.1/ast825/ch9.pdf.
13. B. Keel, *Alternate approaches and the redshift controversy* (2009). Available at https://pages.astronomy.ua.edu/keel/galaxies/arp.html.
14. T. F. Adams, Astrophys. J. Suppl. Ser. **33**, 19–34 (1977).
15. J. J. Condon, S. L. O'Dell, J. J. Puschell, and W. A. Stein, Nature **283**, 357–358 (1980).
16. D. E. Osterbrock and A. T. Koski, Mon. Notices Royal Astron. Soc. **176**, 61–66 (1976).
17. R. Antonucci, Annu. Rev. Astron. Astrophys. **31**, 473–521 (1993).
18. G. Neugebauer, H. J. Habing, R. V. Duinen, H. H. Aumann, B. Baud, C. A. Beichman, D. A. Beintema, N. Boggess, P. E. Clegg, T. De Jong, et al., Astrophys. J. **278**, 1–6 (1984).
19. R. Maiolino and G. H. Rieke, Astrophys. J. **454**, 95–105 (1995).
20. C. Vignall, W. N. Brandt, D. P. Schnieder, G. P. Garmire, and S. Kaspi, Astron. J. **125**, 418–432 (2003).
21. Howard University, *Quasars as cosmic standard candles* (2021). Available at https://www.cfa.harvard.edu/news/quasars-cosmic-standard-candles.
22. NASA, *Hubble space telescope instruments* (2023). Available at https://science.nasa.gov/mission/hubble/observatory/design/instruments.
23. A. Martinazzo, M. Espadoto, and N. Hirata, VISIGRAPP 2020 "Deep learning for astronomical object classification: A case study," in *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020)*, Vol. 5, edited by G. M. Farinella, P. Radeva, and J. Braz (SciTePress, Setúbal, Portugal, 2020), pp. 87–95.
24. N. Busca and C. Balland, "QuasarNET: Human-level spectral classification and redshifting with Deep Neural Networks," preprint arXiv:1808.09955 [astro-ph] (2018).
25. S. Luo, A. P. Leung, C. Y. Hui, and K. L. Li, Mon. Notices Royal Astron. Soc. **492**, 5377–5390 (2020).
26. A. Torfi, *Nearest neighbor classifier – from theory to practice* (2020). Available at https://medium.com/machine-learning-mindset/nearest-neighbor-classifier-a-working-example-c9b0b3103e4e.
27. S. Falocco, F. J. Carrera, and P. Véron, Mon. Notices Royal Astron. Soc. **510**, 161–176 (2021).
28. P. Véron and M. P. Véron-Cetty, Astron. Astrophys. **518** (2010).