Detecting Active Black Holes Using Radial Light Profile Analysis

Alex Fay
Computer Science, Pomona College

April 2020

1 Purpose

Black holes are believed to be at the center of most galaxies. However, detecting active and super massive black holes for study has been historically difficult as previous techniques involved watching star movement and analyzing gravitational forces. As telescope time is incredibly valuable and expensive, this takes time away from analyzing other celestial objects and is incredibly computational and monetarily expensive. Through analyzing galaxy surface radial light profiles, often referred to as Sersic profiles, I aim to build a small archive of potential active, likely super massive black holes for further study.

2 Theory

As matter falls into a black hole, matter becomes ripped apart, releasing massive amounts of energy. Additionally, as matter orbits the black hole, it gains gravitational energy which causes the object to reach high speeds and temperature, releasing additional energy. This causes the surrounding area to shine. In addition, super massive black holes when surrounded by large amounts of super heated gas, will have extremely bright accretion disks with two jets shining from each pole. This is known as a quasar2. Black holes are believed to have significant affect on the structure and geometry of the galaxy causing a central bulge and powering Seyfert galaxies, while the structure of the galaxy equally affects the black hole back. Barred Spiral galaxies funnel gas inward, triggering an active galactic nucleus1. Likewise, ringed galaxies are thought to form from bar collapse. Finally, it is believed that solar wind from black holes causes a massive increase in star formation in the surrounding galactic core. These components when added together help us distinguish black holes according to core brightness and galaxy's shape. In the near future, I will be addressing geometry of rings and bars more thoroughly. In order to detect supermassive, active black holes, I detect high light abnormalities in the galactic core, indicating one or more of these features are present.

Every galaxy has a unique Sersic profile with galaxies of each type having similar trends. Large elliptical galaxies tend to have steep central curves, while lenticular are less bright. By separating each galaxy into different types and calculating the angle of the galaxy from Hubble's camera, I split the galaxies into similar radial subcategories. This will allow for easier detection of abnormalities. Then, using galaxies which are 1.3 times above the expected values, I classify them as potential black hole candidates.

OverView:
1. Collect data from the SDSS archive. This data will be used for finding galaxy images.
2. Using the CSV from SDSS, use the image scraper that I built to retrieve each image.
3. Build the training data by combining voting data from GalaxyZoo. Retrieve Images.
4. Crop each image to desired amount, use CNN to classify each galaxy by type.
5. Using classifications, run data in CNN and place each galaxy in folder by type.
6. Calculate the degree of tilt for each galaxy
7. Develop the radial light profile
8. In the resulting CSV, train a polynomial regression curve by each galaxy type and angle
9. Galaxies with a core 1.3 times brighter than expected willl be classified as likely having an active, supermassive black hole.

3 Collecting Data

The following section is kept from my midterm project and may be skipped if desired.

The Hubble Space Catalog is a massive archive of all Hubble images. As part of Hubble's Mission, a group of researchers have decided to create a 3D virtual map - piecing together each Hubble image into a coherent map. This is called the Sloan Data Sky Survey. This is one of the most widely used, well documented, and reliable data sets within the astronomical community.

For retrieving data, go to skyserver.sdss.org and click on Data and then SQL. Run the SQL query below:

SELECT objID, ra, dec
FROM Galaxy
WHERE
(r - $extinction_r$) between 13 and 24

This will return a CSV of galaxy names and location results which will be used later to retrieve each picture.

Luminosity Intensity is sorted on a log scale, where 1 is extremely bright and 30 is not visible. For spiral galaxies, the average light brightness is 20.2. I took a normal distribution sample in order to max out the number of spiral galaxies, which is believed to have more supermassive black holes than other types of galaxies. Changing the brightness numbers above, I got 34 percent of my data points from 19-20 and 20-21, 13.5 percent from 18-19, 21-22, etc. However, after analyzing these images, many in the 20 and above range were way to faint to work with. Elliptical galaxies have an average intensity of about 15. I decided to take two samples, a normal distribution of spiral and elliptical galaxies. The total images I pulled was 650,000. This returns a csv file with the galaxy names. I have uploaded the csv files into the Data folder on Github.

SDSS has a search image function but it can only handle 100 images at a time. This was far too slow for my purposes so I built my own image search script. Please see imageRetrieval.py in the "Getting Data" Folder. This retrieves each image as a jpg, 120 by 120 pixels (as zoomed in as I can with the galaxy) and labels the galaxy by Brightness and GalaxyID.

4 Galaxy Classification Model

Building Training Data

The GalaxyZoo is a project developed by the SDSS group to help classify galaxies and other celestial objects. Through volunteers, members vote on what type of galaxy and features are present. By averaging the votes on each image, this number is then placed in a chart as a probability of having a certain characteristic. Each image receives 100 votes for a total of 144,000 images. Using this data from the GalaxyZoo, I combined each of the tables into one data set and followed the Hubble classification system by feature to determine the type of galaxy. I built a classifier which runs a series of tests, following Hubble's System, to determine each type of galaxy. Please see DecisionTreeHubbleClassification.jpg under the classification folder for an overview of the Hubble Classification model. The program which runs these tests is galaxyZooTrainingRules.py. This program tests for the features presents and sorts each galaxy. I made a few changes to the classification for my research specific intent. Instead of having 7 elliptical galaxies, I grouped them into their subcategories of Elliptical E0-E5 and Cigar Elliptical E6-E7. The same was applied to spirals as they were separated into bar and noBar subcategories. The final categories used were as follows: Elliptical, Elliptical Cigar, Barred Spiral, Spiral, Lenticular, and Star. The output is a CSV with each image name and classification.

CNN Model
In order to have as much consistency across the dataset as possible, first crop the training images to 212. This is the most similar I can get without losing some of the galaxy. The data we scraped is 120 by 120. By cropping the image, excess noise such as stars will be eliminated and help produce a clean profile.
Using a convectional neural net, sort each image using the training data and SDSS data. A CNN was used due to their ability to pick out features in each image, as each node in the network can be tailored to maximize the reward. With the wide variety of data, CNN's can easily classify multiple classes of data. They typically require massive amounts of data, but this is not a problem as the SDSS has over 6 million images. However CNNs are slow. This takes about 4 hours to train alone.
I used a convectional block with 5 activation layers and relu which takes the returns 0 for negative numbers and x for any positive input x. I experimented with several different activation and this one was both fast and accurate. I added a Dense layer at the end of my $VGG_16()$ model using a sigmoid activation. The final model has a 78 percent accuracy which was tested on 25 percent of the training data.

5 Calculating tilt from Hubble's Camera

Most galaxies, except elliptical galaxies, when viewed face on appear to be circular. However, from a tilted angle, each galaxy becomes an ellipse. By calculating how flattened each circle is, the tilt can be calculated.
In order to do this, start with he equation of a circle. When the object is tilted along the y direction, the resulting radius will become $R^2 * cos^2$. Using Isotropic expansion ranging greater or equal to 1 due to the stretch of the ellipse, solve for the following:

$$\frac{x^2}{f^2 * R^2} + \frac{y^2}{f^2 * R^2 * cos^2\theta} = 1$$

f$^2 = \frac{x^2 cos^2\theta + y^2}{R^2 * cos^2 * \theta}$(1)Plugging back into the original equation we get:

$$\frac{x^2 * cos^2\theta}{x^2 * cos^2\theta + y^2} + \frac{y^2}{x^2 * cos^2\theta + y^2} = 1 \tag{2}$$

Solving for theta this becomes:

$$\theta = cos^{-1}(\frac{-\sqrt{y^2 - b^2}}{\sqrt{a^2 - x^2}}) \tag{3}$$

Such that a and b are the outside radius points of x and y respectively.
In order to do this the following must be found: x radius, y radius, origin. Included within the radial light profile program, is a subsection which finds the

tilt of each image. This is done by first finding the brightest area of the image which is then set as the origin. To find the radius, find the last brightest pixel area in one direction, then do the same 90 degrees from this position. Using this, plug into the equation for theta found above. This gives us our final answer.

6 Radial Light Profile

Using the classified images, run them through the radial profile. This will return a csv with the image name and the radial profile for each image which is used to detect abnormalities. Keep each type of galaxy separate and run separately. For code, refer to radialLight.py

Code Explanation:
1. Assign the center of the 120 by 120 image
2. Shape image and read in image array
3. Use the first two values of each tuple to calculate the radius. We ignore the third as it is the color values.
4. Use binvalue to take the integral of the brightness. This adds each pixel value within the radius. Using r.ravel() is equivelant to dr or slowly incramenting the radius. 5. The radial light profile is the the total light profile (tbin) divided by the increase in radius (nr)
6. return radius
7. Plot using matplotLib

Please see SpiralGalaxyRadialLightProfile.png for an example of a galaxy. You can run this on any image retrieved from sdss, or any image in general. The peak represents the center of the galaxy while the tail represents background noise of empty space. The small dip near the end is from a star in the picture used.

7 Regression

Using the radial profile CSVs, I split the galaxies according to theta, as galaxies that are tilted farther away have less recieved light from Hubble's camera and thus less intense raidal profiles.

Using the csv, set y to the radius in pixels (300 points) and x to the values of intensity. This is trained on 75 percent of the data and uses polynomial regression with 5 degrees. Train on each type of galacy and subclass angles of 1-10, 10 - 30, 30-45. There is significant difference in the first 10 degrees of angle.
After training, select the last 5 points of polynomial line and set as the threshold for galaxy light. if the last 5 points of the profile (on average) are 1.3 times

brighter than this then the galaxy is classified as a black hole.

8 Results

After classifying images, about 5,000 potential candidates were identified out of the 650,000 data set. These images included many ringed galaxies, bar, and galaxies with faint but visible light poles. These are all good indications of active and super massive black holes. Since finding black holes is still a new area of research, there is no clear way to validate this data set except through these features.

9 Next Steps

This summer, I plan to focus on the geometry of galaxies, specifically looking for polar jets to identify quasars. Additionally, I plan to identify galaxies types that have been known to have more active black holes. I also hope to improve my fitting functions to make sure they don't over fit for outliers.

10 Links and References

Retrieving Data using SQL Query Documentation:
http://skyserver.sdss.org/dr16/en/help/docs/realquery.aspxgalview

Image Retrieval Tool:
http://skyserver.sdss.org/dr16/en/tools/chart/listinfo.aspx

Galaxy Zoo/ Hubble Galaxy Classification Tree:
https://data.galaxyzoo.org/gz$_{t}rees/gz_{t}rees.html$

Getting Data From Galaxy Zoo:
https://data.galaxyzoo.org/

Why finding black holes is hard:
https://www.space.com/3457-tricky-task-detecting-black-holes.html

11 Bibliography

1. Goulding, A. D., Matthaey, E., Greene, J. E. (2017). Feeding Black Holes Through Galactic Bars. American Astronomical Society, 6. Retrieved from https://aasnova.org/2017/11/07/feeding-black-holes-through-galactic-bars/

2. Wu, Xue-Bing; et al. (2015). "An ultraluminous quasar with a twelve-billion-solar-mass black hole at redshift 6.30". Nature. 518 (7540): 512–515. arXiv:1502.07418. Bibcode:2015Natur.518..512W. doi:10.1038/nature14241. PMID 25719667.

12 Thankyou

Thankyou to Professor Quintin at Pomona College for giving me this idea and helping me figure out next steps through several meetings. I would not have gotten this far without you.

Thankyou for reading this paper and getting this far. To Professor Gu, thankyou for teaching this class. I think it is incredibly interesting and I'm now considering going to graduate school to do research in this field.