Machine learning for radio galaxy morphology analysis

Mostert, R.I.J.

English summary

Celestial objects have always influenced our human lives.[88] The rising and setting of the sun determine the rhythm of our day, we get tired when the sun sets and wake up when the sun rises. We can navigate by the stars in the sky. We fish to the rhythm of the tides. We grow the crops we eat and adjust our activities to the rhythm of the seasons. We have been monitoring the positions and movements of celestial bodies for millennia, in part because of this.

Because of the immense distances between celestial bodies, it is difficult to observe anything other than the movements of celestial bodies with the naked eye. This first changed with the advent of telescopes in the seventeenth century. With telescopes, we capture much more light than with our eyes, allowing us to map much fainter light sources with greater detail. Celestial objects that looked like nebulae to the naked eye turned out to be entire galaxies when viewed through a telescope: huge collections of stars located at great distances from the Milky Way, our own star system.

A groundbreaking discovery in the nineteenth century revealed that all kinds of different forms of radiation — X-ray, microwave, visible light, infrared, radio — are different manifestations of the same phenomenon: electromagnetic radiation, with the wavelength of the electromagnetic radiation determining the character of the radiation. A second important discovery is that atoms and molecules absorb and emit light, each with their own characteristic set of wavelengths. Telescopes, in combination with techniques to separate light into its constituent wavelengths, thus enables us to study the composition of celestial objects. A third important discovery is that by measuring the shift of the atomic and molecular spectra of celestial objects, we can also determine the distance to these objects, even if they are far away from us. This so-called 'redshift' of spectral lines is most easily detected in visible or infrared light.

A fourth important realization is that with different wavelengths, different physical processes are operating and the transparency and reflectivity of materials are wavelength dependent. In other words, if you observe the world with light of different wavelengths, the world looks different. With an infrared camera, we mainly see objects that radiate heat. X-rays pass through skin and muscle, but are more readily absorbed by bones. Similarly, the universe looks different at different wavelengths.

By chance, an engineer that looked into the source of noise on telephone lines in 1932 discovered that, in addition to visible light, radio waves from extra-terrestrial origins also descend upon Earth. Still, it took years before the origin of these radio waves could be accurately determined. In general, the resolution of a telescope scales with the wavelength of light divided by the diameter of the telescope. The larger the diameter of the telescope the more details are visible. The larger the wavelength of the light, the fewer details are visible. A low frequency radio telescope requires a diameter of four thousand kilometres to achieve the same resolution as a one-meter optical telescope. With radio interferometry, invented in the 1950s, multiple individual radio antennas can be combined to create one big radio telescope, where the resolution scales with the wavelength divided by the greatest distance between two antennas. In this way, it is possible to reconstruct a very detailed image of the sky even with radio waves.

With visible light, the night sky is mostly filled with stars. With radio waves, we mostly see the remains of exploded stars, called supernovae, and the cores of galaxies. There, in the cores of galaxies,

[88] In high school, we learn that evolutionary pressure is created by the climate and the flora and fauna that surrounds us, but we do not often consider the direct evolutionary pressure of the Sun, the nearest star. It is no accident that our eyes are most sensitive to the colours of light emitted by the Sun with the greatest intensity.

supermassive black holes appear to reside. These black holes often produce strong magnetic fields, and under the influence of the black hole's gravity, a disk of matter (the 'accretion disk') often forms around the black hole. The accretion disk emits (UV) radiation, and along the magnetic fields, electrically charged particles can be accelerated, emitting radio radiation in the process.

The activity of these black holes affects the galaxy they inhabit. Both the radiation from the accretion disk and the accelerated electrically charged particles affect the rate at which new stars are formed in the galaxy in question. The jet of charged particles, from the black hole's magnetic field, can for example blow away or expel gas from the galaxy, preventing the formation of stars from this gas, and (sometimes) reducing the amount of gas available to feed the black hole.

Since light travels at a finite speed and celestial bodies are very far away from us, it takes a significant amount of time for light from celestial bodies to reach us. Even the nearest star, the Sun, appears to us on Earth as it was eight minutes ago.[89] The further a celestial body is removed from us, the further we look back in time, just as archaeologists go back in time with each successive uncovered layer of earth. By observing radio waves, we see activity from black holes that are millions or even billions of light years away from us, which means we can look billions of years back in time.

By classifying the different types of black holes at different distances from us and recording how strong the activity of each of these populations is, we gain insight into the evolution of black holes, star systems and the universe over time. We gain this insight directly, by looking at trends from observations, but also indirectly by being able to test theoretical and numerical models of the universe for consistency with the analysis resulting from observations.

To find large numbers of black holes (and therefore galaxies) both near earth and far away, we need a large radio telescope observing large swaths of the sky. Unknown to many, the northeast of the Netherlands hosts the inconspicuous centre of Europe's largest (radio) telescope. The so-called 'LOw Frequency ARray', or LOFAR for short, is a modular radio telescope consisting of hundreds of antennas from one to two meters in size. These antennas are set up in countries as far apart as Ireland and Latvia in the longitudinal direction, and from Sweden to France in the latitudinal direction, and are interconnected by fibre optic cables. By cleverly combining the signals captured by these antennas, images can be made with a resolution that does not depend on the size of a single antenna, but on the greatest distance between antennas. With LOFAR's Dutch antennas, an international consortium of astronomers is making observations of the entire northern sky. These observations, under the name 'LOFAR Two-metre Sky Survey', or LoTSS for short, capture millions of celestial objects.

Hundreds of thousands of these celestial objects in this map appear large enough, to say something about their shape. The shape or pattern emerging from the radiation from the electrically charged particles emanating from an active supermassive black hole provides (indirect) insight into the strength of the black hole's activity and the immediate environment of the galaxy in which the black hole is located. However, manually classifying hundreds of thousands of objects based on their shape is extremely time- consuming. We can only unlock the full potential of the terabytes of collected data, if we can automate the morphological classification process and other analysis steps.

Classical ways of applying automatic classification based on shape require devising and coding a set of rules that can separate the expected patterns/shapes into pre-conceived classes. However, shapes and brightness gradients in images are notoriously difficult to capture quantitatively in language or computer code. The alternatives we apply in this thesis are various forms of 'machine

[89]"Je kijkt terug in de tijd, als je naar haar kijkt" is the accompanying quote from the song "Sterrenstof" by De jeugd van tegenwoordig.

learning'. Machine learning is a form of artificial intelligence in computer science that includes a wide range of algorithms that 'learn' to distinguish patterns in data sets. As such, we can obtain quantitative models of the morphologies of radio sources in radio maps that are not based on self-imposed prescriptions but on a large number of examples from our data. Such a set of examples for a machine learning algorithm is called a 'training dataset'.

Chapter 1 of this thesis paints a picture of the rise of radio astronomy, describes the challenges of automatic image recognition (also known as 'computer vision'), and introduces the machine learning algorithms we use in the following chapters. The following is a summary of these subsequent chapters.

Chapter 2. Revealing rare morphologies in LoTSS with self-organizing maps.
Never before have we captured the radio emission from so many active black holes in such detail as with the radio maps of LoTSS. From previous observations of active black holes, we already have an idea of the usual shapes that radio sources manifest at 150 MHz. However, many discoveries in science are serendipitous and unforeseen phenomena can be discovered by investigating outlying data points. Therefore, in this chapter, we look for the active black holes with morphologies that are least common in our data. With unsupervised learning, a form of machine learning, we can cluster our data based on morphology without imparting our expectations about expected/usual shapes. Specifically, we cluster our data using the 'self-organizing map' algorithm. With this algorithm, we place each radio source at a particular spot in a 2D space, where radio sources that are morphologically more similar are closer together ('clustered'). The algorithm places radio sources with rare morphology far away from all other radio sources within the 2D space. In this way, we can identify radio sources with rare morphologies. In our data, we find a wide range of extra-galactic radio sources: (strongly curved) fountains of radio emission from active black holes, diffuse emission from clustered galaxies (so-called 'cluster haloes'), fossil emission from (formerly) active black holes, and radio emission from star-forming (spiral) galaxies.

Chapter 3. Radio source components association with region-based convolutional neural networks.
The magnetic field of active black holes accelerates electrically charged particles along two diametrically positioned cones. These two 'particle fountains' emit radio waves, but this radiation is not equally bright everywhere along the fountain. It often happens that in certain places the radiation from the two fountains do not exceed the measured radio background noise. Nevertheless, the shape of the two fountains often shows that they belong together. Using neural networks, a form of machine learning, we teach our model which radio emission regions (in other words, which components of a radio source) belong together and which are unrelated. As a training dataset, we use 6, 158 radio source components that were manually annotated by astronomers of the LOFAR collaboration. With the arrival of more radio sources in the course of the LoTSS project, manual annotation by astronomers was too time consuming and large-scale public volunteering was called upon. For apparently large and bright active black holes, our neural network achieves a result similar to that attained by the manual volunteering solution. As such, this work automates a part of the complex process from converting radio images to a sky-catalogue of active black holes.

Chapter 4. Finding candidates of black hole remnants with the help of machine learning.

Supermassive black holes at the centre of galaxies are not always active. Sometimes the flow of electrically charged particles from a supermassive black hole that was previously active stops. The radiation from the previously accelerated charged particles at a wavelength of 150 MHz still 'glows' for a long time. Based on the morphology of this fossil radio emission, we can sometimes assume or suspect that we are dealing with the remnants of an active black hole. In this chapter, we look for a way to reduce the manual inspection required to find candidate black hole remnants in LoTSS. To make the classification process repeatable and quantitative, we wrote a code that extracts a wide range of features from each radio source. One of the morphological features we extract is the location of a radio source in a self-organizing map (thereby, learning from our experience from Chapter 2). After the feature extraction, we use an algorithm that takes all the extracted features into account to classify whether a radio source is a candidate black hole remnant. Based on the small number of remnant examples at our disposal, we managed to create a classifier that reduces the number of radio sources larger than 1 arcmin that are require visual inspection to check whether they are likely remnant candidates by 75%. This result is a step in the right direction, but full automation has not yet been achieved for finding all black hole remnant candidates.

Chapter 5. Finding giant active black holes with machine learning and estimating the properties of the intrinsic giant population with forward modelling.

One of the enigmas of the universe is that the universe is magnetized. Several theories try to explain this phenomenon, and one of them states that active black holes contribute significantly to the magnetization of the universe. Presently, the volume covered by the magnetic cones emanating from the active black holes observed so far is relatively small. However, there is also a class of active black holes known as giants. Active black holes are labelled giants if they extent at least 0.7 megaparsec (2.28 million light-years). Observationally, relatively few giants are known, but if the intrinsic population of giants were large, black holes could be a significant factor in the magnetization of the universe. To investigate this, we built a pipeline to search for candidate giants in LoTSS. This pipeline consists of an aggregation of five different methods from the literature including the methods developed in Chapter 3. After searching for candidates, we manually verified all candidate giants in LoTSS, identifying over eight thousand new giants. Combined with previously found giants from the literature, this puts the number of observed giants above ten thousand for the first time. Based on our giants, the giants from the literature, and a forward model, we estimate the intrinsic length distribution of giants and the number of giants per unit volume. Using these numbers, we postulate that the density of the intrinsic number density of giants is consistent with the intrinsic number density of 'ordinary' active black holes. In addition, we estimate that the volume covered by the intrinsic giant population may indeed contribute significantly to the magnetization of the universe.