



Universiteit
Leiden
The Netherlands

Machine learning for radio galaxy morphology analysis

Mostert, R.I.J.

Citation

Mostert, R. I. J. (2024, January 25). *Machine learning for radio galaxy morphology analysis*. Retrieved from <https://hdl.handle.net/1887/3715061>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3715061>

Note: To cite this publication please use the final published version (if applicable).

Nederlandstalige samenvatting

Hemelobjecten hebben onze mensenlevens altijd beïnvloed.⁹⁰ De opkomst en ondergang van de zon bepaalt het ritme van onze dag, we worden moe als de zon ondergaat en worden wakker als de zon opkomt. We kunnen navigeren aan de hand van de sterren aan de hemel. We vissen op het ritme van de getijden. We verbouwen de gewassen die we eten en passen onze activiteiten aan aan het ritme van de seizoenen. De stand en beweging van hemellichamen houden we mede daarom al millennia in de gaten.

Vanwege de immense afstanden tussen hemellichamen, is het lastig om met het blote oog iets anders dan de bewegingen van de hemellichamen in de gaten te houden. Met de komst van telescopen in de zeventiende eeuw kwam daar voor het eerst verandering in. Met telescopen vangen we veel meer licht op dan met onze ogen, waardoor we veel zwakkere lichtbronnen met meer detail in kaart kunnen brengen. Hemelobjecten die er met het blote oog uitzagen als nevels, bleken door een telescoop bekeken hele sterrenstelsels te zijn: enorme verzamelingen aan sterren die op grote afstand van de Melkweg, ons eigen sterrenstelsel, staan.

Een baanbrekende ontdekking in de negentiende eeuw onthulde dat allerlei verschillende vormen van straling — röntgenstraling, microgolfstraling, zichtbaar licht, infrarood, radiostraling — verschillende verschijningsvormen zijn van een en hetzelfde fenomeen: elektromagnetische straling, waarbij de golflengte van de elektromagnetische straling het karakter van de straling bepaalt. Een tweede belangrijke ontdekking is dat atomen en moleculen licht absorberen en uitzenden met elk een eigen karakteristieke set golflengtes. Telescopen, in combinatie met technieken om licht op te breken in verschillende golflengtes maken het dus mogelijk om de samenstelling van hemelobjecten te onderzoeken. Een derde belangrijke ontdekking, is dat we met het verschuiven van atomaire en moleculaire spectra van hemelobjecten ook de afstand tot deze objecten kunnen bepalen, zelfs als ze ver van ons af staan. Deze zogenaemde ‘roodverschuiving’ van het lichtspectrum is het makkelijkst te detecteren in zichtbaar of infrarood licht.

Een vierde belangrijke realisatie is dat met verschillende golflengtes, verschillende fysische processen zichtbaar zijn en dat de transparantie en reflectie van materialen golflengte afhankelijk zijn. Anders gezegd, als je de wereld bekijkt met verschillende golflengtes ziet de wereld er anders uit. Met een infraroodcamera zien we voornamelijk objecten die warmte uitstralen. Röntgenstralen gaan door huid en spieren heen, maar komen moeilijker door bot heen. Ook de hemel ziet er bij andere golflengtes anders uit.

Bij toeval ontdekte een onderzoeker die in 1932 op zoek was naar de bron van ruis op telefoonlijnen, dat er behalve zichtbaar licht, ook radiogolven uit de ruimte neerdalen op aarde. Toch duurde het nog jaren voordat de oorsprong van deze radiogolven nauwkeurig kon worden bepaald. In het algemeen geldt dat de resolutie van een telescoop schaal met de golflengte van het licht gedeeld door de diameter van de telescoop. Hoe groter de diameter van de telescoop hoe meer details zichtbaar zijn. Hoe groter de golflengte van het licht, hoe minder details zichtbaar zijn. Een laagfrequentie radiotelescoop moet een diameter hebben van vierduizend kilometer om dezelfde resolutie te behalen als een optische telescoop van één meter. Met radio interferometrie, bedacht in de jaren vijftig van de vorige eeuw, kunnen meerdere radioantennes gecombineerd worden tot één radiotelescoop,

⁹⁰We leren op de middelbare school dat er evolutionaire druk uit gaat van het klimaat en de ons omringende flora en fauna, maar we staan niet vaak stil bij de evolutionaire druk van de Zon, de dichtstbijzijnde ster. Onze ogen zijn niet toevallig het meest gevoelig voor de kleuren licht die de Zon met de grootste intensiteit uitstraalt.

waarbij de resolutie schaal met de golflengte gedeeld door de grootste afstand tussen twee antennes. Op deze manier is het mogelijk om ook met radiogolven een zeer gedetailleerd beeld van de hemel te reconstrueren.

Met zichtbaar licht zien we aan de hemel vooral sterren. Met radiogolven zien we voornamelijk de overblijfselen van ontplofte sterren, supernovae geheten, en de kernen van sterrenstelsels oplichten. Daar, in de kernen van sterrenstelsels, blijken superzware zwarte gaten te huizen. Deze zwarte gaten produceren vaak sterke magneetvelden en onder invloed van de zwaartekracht van de zwarte gaten vormt zich vaak een schijf van materie (de ‘accretieschijf’) om het zwarte gat. De accretieschijf zendt (UV-)straling uit, en langs de magneetvelden kunnen elektrisch geladen deeltjes versneld worden die daarbij radiostraling uitzenden.

De activiteit van deze zwarte gaten hebben invloed op het sterrenstelsel waar ze in huizen. Zowel de straling van de accretieschijf als de versnelde elektrisch geladen deeltjes oefenen invloed uit op het tempo waarop nieuwe sterren geboren worden in het desbetreffende sterrenstelsel. Zo kan de straal van geladen deeltjes, vanuit het magnetisch veld van het zwarte gat, gas wegblazen uit het sterrenstelsel, waardoor er uit dit gas geen sterren meer gevormd kunnen worden, en er (soms) ook minder gas is om het zwarte gat te voeden.

Aangezien licht met een eindige snelheid reist en hemellichamen erg ver van ons af staan kost het een significante hoeveelheid tijd voordat het licht van hemellichamen ons bereikt. Zelfs de dichtstbijzijnde ster, de Zon, zien we altijd zoals ze er acht minuten geleden uit zag.⁹¹ Hoe verder een hemellichaam van ons af staat, hoe verder we terugkijken in de tijd, zoals archeologen terug in de tijd gaan met elke volgende aardlaag. Met radiostralen zien we activiteit van zwarte gaten die miljoenen of zelfs miljarden lichtjaren van ons af staan en daarmee kijken we dus ook miljarden jaren terug in de tijd.

Door de verschillende soorten zwarte gaten op verschillende afstanden van ons vandaan te classificeren en vast te leggen hoe sterk de activiteit van elk van deze populaties is, krijgen we inzicht in de evolutie van zwarte gaten, sterrenstelsels en het Universum over de tijd. Dit inzicht krijgen we direct, door te kijken naar de trends van observaties, maar ook indirect doordat we theoretische en numerieke modellen van het Universum kunnen toetsen op consistentie met de analyse die voortkomt uit observaties.

Om grote aantallen zwarte gaten (en daarmee sterrenstelsels) te vinden die zowel dichtbij als ver weg staan, hebben we een grote radiotelescoop nodig die grote stukken van de hemel observeert. In het Noordoosten van Nederland staat het onopvallende centrum van de grootste (radio)telescoop van Europa. De zogeheten “LOW Frequency ARray”, kortweg LOFAR, is een modulaire radiotelescoop die bestaat uit honderden antennes van een tot twee meter groot. Deze antennes staan opgesteld van Ierland tot Letland, en van Zweden tot Frankrijk en zijn met glasvezel met elkaar verbonden. Door de signalen die deze antennes opvangen op een slimme manier te combineren kunnen er afbeeldingen gemaakt worden met een resolutie die niet afhangt van de grootte van een enkele antenne, maar van de grootste afstand tussen de antennes. Met de Nederlandse antennes van LOFAR maakt een internationaal consortium van sterrenkundigen observaties van de gehele noordelijke hemel. Deze observaties, onder de noemer ‘LOFAR Two-metre Sky Survey’ kortweg LoTSS, brengen miljoenen hemelobjecten in kaart.

Honderd duizenden van deze hemelobjecten in deze kaart staan dichtbij genoeg, of zijn groot genoeg, om iets te kunnen zeggen over hun vorm. De vorm of het patroon dat de straling van de elektrisch geladen deeltjes die van een actief superzwaar zwart gat komen aannemen, geven (indirect)

⁹¹“Je kijkt terug in de tijd, als je naar haar kijkt” is het bijbehorende citaat uit ‘Sterrenstof’ van De Jeugd van Tegenwoordig.

inzicht in de kracht van de activiteit van het zwart gat, en inzicht in de directe omgeving van het sterrenstelsel waar het zwarte gat zich in bevind. Honderd duizenden objecten handmatig classificeren op basis van hun vorm is echter extreem tijdrovend. Pas als we het morfologische classificatieproces, en andere analyse-stappen, kunnen automatiseren kunnen we het volledig potentieel van de terabytes aan verzamelde observaties ontsluiten.

Klassieke manieren om automatische classificatie op basis van vorm toe te passen vereisen het bedenken en coderen van een set regels die de verwachte patronen van elkaar kan scheiden in de vooraf bedachte klassen. Vormen en helderheid gradiënten in afbeeldingen zijn echter notoir lastig kwantitatief vast te leggen in taal of computercode. Een alternatief dat we in deze thesis toepassen zijn verschillende vormen van machinaal leren (ook wel “machine learning”). Machine learning is een vorm van kunstmatige intelligentie in de informatica die een breed scala aan algoritmes omvat die zelf ‘leren’ van patronen in data sets. Op deze manier kunnen we softwarematig een model verkrijgen van de morfologieën van radiobronnen in de radiokaarten, niet op basis van zelfbedachte voorschriften, maar op basis van een groot aantal voorbeelden uit onze data — zo’n set aan voorbeelden voor een machine learning algoritme noemen we een ‘training dataset’.

Hoofdstuk 1 van deze thesis schetst een beeld van de opkomst van de radiosterrenkunde, benoemd de uitdagingen van automatische beeldherkenning (ook wel ‘computer vision’) en introduceert de machine learning algoritmes die we gebruiken in de volgende hoofdstukken. Hieronder volgt een samenvatting van deze volgende hoofdstukken.

Hoofdstuk 2. Zeldzame morfologieën in LoTSS onthullen met self-organising maps.

Niet eerder hebben we van zoveel actieve zwarte gaten de radiostraling zo scherp in beeld gehad als met de radio-kaarten van LoTSS. Van eerdere observaties van actieve zwarte gaten hebben we al een idee van de gebruikelijke vormen die de radiostraling op 150 MHz aanneemt. Nieuwe ontdekkingen in de wetenschap liggen soms echter verscholen op plekken waar we niet eerder gezocht hebben, of achter fenomenen die juist niet passen in ons huidige theorieën of verwachtingen. In dit hoofdstuk zoeken we daarom naar de actieve zwarte gaten met morfologieën die het minst vaak voorkomen in de data. Met ‘unsupervised learning’, een vorm van machinaal leren, kunnen we onze data clusteren op basis van morfologie zonder onze verwachtingen over gebruikelijke vormen mee te geven. Specifiek clusteren we onze data met behulp van het ‘self-organizing map’ algoritme. Daarmee plaatsen we elke radiobron op een bepaalde plek in een 2D ruimte, waarbij radiobronnen die morfologisch meer op elkaar lijken, dichter bij elkaar staan (‘geclusterd’ zijn). Radiobronnen met een zeldzame morfologie plaatst het algoritme ver van alle andere radiobronnen binnen de 2D ruimte. Op deze manier kunnen we radiobronnen met zeldzame morfologieën identificeren. In onze data vinden we een breed scala aan extragalactische radiobronnen: van (sterk gekromde) deeltjes-fontijnen van actieve zwarte gaten, diffuse emissie van geclusterde sterrenstelsels (zogenoemde ‘cluster haloes’), de fossiele emissie van (voormalig) actieve zwarte gaten, tot radiostraling van stervormende (spiraal)sterrenstelsels.

Hoofdstuk 3. Radiobron componenten associatie met regio-gebaseerde convolutionele neurale netwerken.

Het magneetveld van actieve zwarte gaten versnelt elektrisch geladen deeltjes langs twee diametraal gepositioneerde kegels. Deze twee ‘deeltjes-fonteinen’ zenden radiostraling uit, maar deze straling is niet overal langs de fontein even fel. Het komt vaak voor dat de straling van de twee fonteinen op

bepaalde plekken niet boven de gemeten radio-achtergrond-ruis uitkomt. Toch is aan de vorm van de twee fonteinën vaak wel af te leiden dat ze bij elkaar horen. Met behulp van neurale netwerken, een vorm van machinaal leren, leren we ons model welke radio-emissie (anders gezegd: welke componenten van een radiobron) bij elkaar hoort en welke niet gerelateerd zijn. Als training dataset gebruiken we 6.158 radiobron componenten die door sterrenkundigen van de LOFAR-collaboratie handmatig geannoteerd waren. Met de komst van meer radiobronnen in de loop van het LoTSS-project was handmatige annotatie door sterrenkundigen niet toereikend en werd onder andere gebruikt maakt van grootschalige publieksparticipatie. Voor felle actieve zwarte gaten bereiken we met ons neurale netwerk een resultaat dat vergelijkbaar is met de handmatig publieksparticipatie oplossing. Daarmee automatiseren we een deeltaak in het complexe verwerkingsproces van een radiokaart naar een catalogus van actieve zwarte gaten.

Hoofdstuk 4. Kandidaten van zwart gat restanten vinden met behulp van machine learning.

De superzware zwarte gaten in het centrum van sterrenstelsels zijn niet altijd actief. Soms stopt de stroom van elektrisch geladen deeltjes van een superzwaar zwart gat dat eerder wel actief was. Omdat de straling van de eerder versnelde geladen deeltjes bij een golflengte van 150 MHz nog lang nagloeit kunnen we soms op basis van de morfologie van de radio-emissie stellen of vermoeden dat we te maken hebben met de restanten van een actief zwart gat. In dit hoofdstuk gaan we op zoek naar een manier om het vinden van kandidaat zwarte gat restanten in LoTSS minder arbeidsintensief te maken. Om het classificatieproces herhaalbaar en kwantitatief te maken, schrijven we code die een aantal kenmerken van radiobronnen extraheren. Een van de morfologische kenmerken die we extraheren is de locatie van een radiobron in een self-organizing map waarbij we leren van onze ervaring uit Hoofdstuk 2. Vervolgens gebruiken we een ander algoritme om op basis van de verschillende kenmerken te classificeren of een radiobron een kandidaat zwart gat restant is. Op basis van het kleine aantal voorbeelden dat we tot onze beschikking hebben lukt het ons om bij voorbaat 75% van alle radiobronnen groter dan 1 boogminuut af te wijzen als kandidaat. Daarmee zetten we een stap in de goede richting, maar is volledige automatisering nog niet bereikt voor het vinden van zwarte gat restanten.

Hoofdstuk 5. Reuzen actieve zwarte gaten vinden met machinaal leren en de eigenschappen van de intrinsieke reuzen populatie afschatten met voorwaarts modelleren.

Een van de raadsels van het Universum is dat het Universum gemagnetiseerd is. Er zijn verschillende theorieën die dit fenomeen proberen te verklaren en één daarvan stelt dat actieve zwarte gaten een significante bijdrage leveren aan de magnetisatie van het heelal. Nu is het volume dat bestreken wordt door de magnetische kegels die uit de actieve zwarte gaten komen die men tot nog toe geobserveerd heeft relatief klein. Er is echter ook een klasse actieve zwarte gaten die bekend staat als reuzen. Zwarte gaten krijgen het label ‘reus’ als ze op zijn minst een grootte hebben van 0.7 megaparsec (2.28 miljoen lichtjaar). Observatieel gezien zijn er relatief weinig reuzen bekend, maar als de intrinsieke populatie reuzen groot zou zijn kunnen zwarte gaten een significante factor zijn bij de magnetisatie van het Universum. Om dit te onderzoeken bouwen we een pijplijn om kandidaat reuzen te zoeken in LoTSS. Deze pijplijn bestaat uit een samenvoeging van vijf verschillende methodes uit de literatuur waaronder de methode uit Hoofdstuk 3. Na het zoeken van kandidaten verifiëren handmatig alle kandidaat-reuzen in LoTSS, waarmee we ruim achtduizend nieuwe reuzen identificeren. In combinatie met eerder gevonden reuzen uit de literatuur komt het aantal geob-

serveerde reuzen daarmee voor het eerst boven de tienduizend uit. Op basis van onze reuzen, plus de reuzen uit de literatuur, modelleren we een schatting van de intrinsieke lengte verdeling van reuzen en een schatting van het aantal reuzen per volume-eenheid. Aan de hand van deze getallen opperen we dat de dichtheid van het intrinsiek aantal reuzen strookt met de dichtheid van het intrinsiek aantal 'gewone' actieve zwarte gaten. Daarnaast schatten we dat het volume dat door de intrinsieke reuzen populatie bestreken wordt inderdaad een significante bijdrage kan leveren aan de magnetisatie van het Universum.

