James Dawson

Title

Modelling galaxy emission-line kinematics using self-supervised learning

Abstract

In the upcoming decades large facilities, such as the SKA, will provide resolved observations of the kinematics of millions of galaxies. In order to assist in the timely exploitation of these vast datasets we have explored the use of self-supervised, physics aware neural networks capable of Bayesian kinematic modelling of galaxies. In this talk I will present the network's ability to model the kinematics of cold gas in galaxies with an emphasis on recovering physical parameters and accompanying modelling errors. The models discussed are able to recover rotation curves, inclinations and disc scale lengths for both CO and HI data which match well with those estimated in the literature. The models are also able to provide modelling errors over learned parameters thanks to the application of quasi-Bayesian Monte-Carlo dropout. This work shows the promising use of machine learning and, in particular, self-supervised neural networks in the context of kinematically modelling galaxies observed using interferomers such as ALMA and VLA as well as IFU instruments like SDSS (MaNGA).

James Dawson

36 The Rise, Cardiff, CF14 0RD – UK

+44 7528109006 • dawsonj5@cardiff.ac.uk
www.sites.google.com/view/jamesmdawson • jamesdawson-spacemeerkat
SpaceMeerkat

Education

Cardiff University — **Cardiff**
Doctor of Philosophy(Physics & Astronomy) — *2017–2021*
Using machine learning to increase the impact of ALMA in preparation for the SKA

Cardiff University — **Cardiff**
MSc Astrophysics, Distinction — *2016–2017*
12 month programme including a 3 month academic placement at Cavendish Laboratory

Cardiff University — **Cardiff**
BSc Astrophysics, 2:1 — *2013–2016*
3 Year program mastering Physics and Astronomy through modules and laboratory sessions

Ermysted's Grammar School — **North Yorkshire**
GCSEs & A-levels — *2009–2013*
7 years leading to completing A-level qualifications prior to university

PhD project

title: *Machine learning with ALMA in preparation for the SKA*

location: Cardiff University

supervisors: Dr Timothy Davis & Dr Edward Gomez

description:
Project 1. Refining the extraction of kinematic features of molecular gas in galaxies using machine learning. These models embed kinematic features into lower dimensional hyperparametric space and are used to rapidly differentiate disturbed gas structures from orderly rotating disks. (Paper quick-link)
Project 2. Developing self-supervised Bayesian neural networks for improving kinematic modelling efficiency for the Square Kilometre Array. These networks, built using PyTorch, predict rotation curves of atomic Hydrogen disks in galaxies and are made to compete with current Bayesian tilted ring fitting routines. (Paper quick-link)
Project 3. Pilot testing the use of self-supervised neural networks with MaNGA data release for automated returning of the baryonic Tully Fisher relation. Comparisons for this test will be made with similar publications using SAMI data in order to determine the next steps in applying these methods to Hector data.

Master thesis

title: *Machine learning Insights into the Next-Generation Transit Survey*

location: Cavendish Laboratory - University of Cambridge

supervisors: Professor Didier Queloz & Dr Edward Gillen

description: Investigating feature selection for variable star classification using the Random Forest classifier. In this work I used the Random Forest classifier to rank variable star features in order to improve the automated removal of false-positive exoplanet detections for the NGTS. Feature selection involved the use of Bayesian decomposition of light-curve as well as automated feature extraction routines to compare hand-selected versus automated methods. (copies available upon request)

Work experience

Industrial

AI research lab — **Cheltenham**
Data Scientist — *March–July 2021*
Further using machine learning to predict traffic behaviour in large scale networks
Detailed achievements: Pending

Institute for Mathematical Research — **Bristol**
Data Scientist — *Jun–Sept 2019*
Using machine learning to predict traffic behaviour in large scale networks
Detailed achievements:
- Using convolutional autoencoders to cluster high dimensional data
- Making direct comparisons of machine learning embeddings with UMAP clustering characteristics

Academic

IRAM - Institut de Radioastronomie Millimétrique — **Pico Veleta**
Observer — *July–2019*
5 days of heterodyne pool observing at the 30m IRAM telescope in Spain

Saint Michel - l'Observatoire — **Haute-Provence**
Observer — *April–2017*
5 nights of observing using the 0.8m and 1.2m optical telescopes in southern France

Computational skills

Python coding:
- Advanced python skills including the use of deep learning libraries such as PyTorch, TensorFlow, and Keras
- Application of CUDA GPU accelerated computation for deep learning frameworks
- Jupyter notebook creation for tutorials and exemplar code
- Various packages for handling data including data analysis libraries and plotting routines
- Rudimentary use of Apache Spark for distributed data handling
- Frequent use of multiprocessing capabilities of `multiprocessing` and `MPI`.
- Advanced experience in creating PyPi packages for publicly distributing code

SQL:
- Astronomical simulation database querying and indexing
- Distributed database querying and manipulation

Github:
- Version control while developing astronomical tools
- Used to create tagged package releases
- Experience using hooks between Github and Read The Docs for compiling documentation

Bash scripting:
- Triggering operations in shells -typically Unix commands and Python scripts

LaTeX:
- Scientific journal paper creation
- Fast, efficient, creation of formal documents such as interim reports and scientific journal papers

Operating systems:
- Linux (advanced use)
- Windows (advanced use)

Microsoft Office:
- Proficient in using Word, PowerPoint, Excel, and Outlook

Communication and engagement

Kavli Sponsored meeting *Petabytes to Science* - Boston, Massachusetts (November 2019):

- Presenting machine learning approaches for studying the kinematics of galaxies in the era of the Square Kilometre Array

ECML PKDD - Würzburg, Germany (September 2019):

- Presenting my winning solution for the European Space Agency's Ariel Machine Learning Data Challenge

ALMABO19 - Bologna, Italy (September 2019):

- Poster showcasing the use of machine learning approaches for studying the kinematics of cold gas in galaxies to maximise the impact of the Atacama Large Millimeter/submillimeter Array

Whispers workshop on hyperspectral image and signal analysis - Amsterdam, The Netherlands (September 2019):

- Presenting results showcasing machine learning approaches for working with interferometric data cubes and their place in the *big data* revolution

Machine Learning Applications for Astronomy - Nottingham, UK (September 2018):

- Presenting initial machine learning approaches and avenues for working with interferometric data cubes for the Atacama Large Millimeter/submillimeter Array

University MSc open day - Cardiff, UK (July 2020):

- Engaging with prospective MSc students at Cardiff University at the invitation of existing MSc coordinators.

Prizes and scholarships

Dr Benjamin Glyndwr Owen Scholarship
Cardiff University *2020*

Master's Excellence Scholarship
Cardiff University *2016*

Languages

English: Fluent native speaking, reading, and writing

Russian: A2 level beginner. Basic level speaking, reading, and writing

Motor vehicle qualifications

- I hold a full UK driving license for class B vehicles
- I hold a full Category A UK motorcycle license

Interests

Machine learning challenges: I am the winner of the 2019 worldwide ARIEL exoplanet machine learning challenge and I enjoy taking part in challenges outside my field of research in my free time including the current SKA SDC2 challenge.

Drumming: I play drums in actively performing bands and session for various artists in studio recordings as well as live performances

Swimming: I have a long career history of swimming at national and international level

Publications

[1] Timothy A. Davis, Jenny E. Greene, Chung-Pei Ma, John P. Blakeslee, **James M. Dawson**, Viraj Pandya, Melanie Veale, and Nikki Zabel. The MASSIVE survey - XI. What drives the molecular gas properties of early-type galaxies. *Monthly Notices of the Royal Astronomical Society*, 486(1):1404–1423, Jun 2019.

[2] Timothy A. Davis, Nikki Zabel, and **James M Dawson**. KinMS: Three-dimensional kinematic modelling of arbitrary gas distributions. page ascl:2006.003, June 2020.

[3] Nikolaos Nikolaou, Ingo P. Waldmann, and **James M Dawson** et al. Lessons Learned from the 1st ARIEL Machine Learning Challenge: Correcting Transiting Exoplanet Light Curves for Stellar Spots. *ApJ*, 2020 - submitted.

[4] **James M Dawson**, Timothy A Davis, Edward L Gomez, and Justus Schock. A Self-supervised, Physics-aware, Bayesian Neural Network architecture for modelling galaxy kinematics in preparation for the SKA. *MNRAS*, 2020 - submitted.

[5] **James M Dawson**, Timothy A Davis, Edward L Gomez, Justus Schock, Nikki Zabel, and Thomas G Williams. Using machine learning to study the kinematics of cold gas in galaxies. *Monthly Notices of the Royal Astronomical Society*, 491(2):2506–2519, 11 2019.