Henry Wilde

PhD thesis

Title
: **New methods for algorithm evaluation and cluster initialisation with applications to healthcare**

Supervisors
: Dr Jonathan Gillard, Dr Vincent Knight, Mr Kendal Smith (NHS Wales)

Description
: This thesis offers a thorough utilisation of machine learning techniques to understand variation in healthcare data; this is achieved by assessing, creating and applying machine learning techniques in a novel manner. The results of this are new perspectives on algorithm evaluation and fair clustering, as well as attaining actionable insights into a healthcare population using routinely gathered, administrative datasets. In addition to these contributions, this thesis is accompanied by a collection of well-developed research software packages.

Education

2017–2021
: **PhD Applied Statistics, OR and Data Analytics**, *Cardiff University*.
  - *Research interests:* healthcare modelling, clustering, algorithm evaluation, game theory, research software development.
  - *Key skills:* confidence in the principles of data science and mathematical programming, ability to develop industrial relationships, proficiency in LaTeX and Python.

2014–2017
: **BSc Mathematics (First Class Honours)**, *Cardiff University*.
  - *Key areas of study:* methods for data mining, game theory, algorithms and heuristics.
  - *Projects included:* building a simulation of a hospital emergency department, and a principal analysis of two game-theoretic strategies within an Iterated Prisoner's Dilemma tournament.

Relevant experience

Feb 2021–present
: **Research associate**, *Cardiff University*.
  I am a member of the mathematical modelling team investigating the potential for wastewater sampling in disease prediction, and particularly COVID-19. This role is split between two major projects: the WeWASH project with Welsh Government, and an international project with the University of Campinas funded by the Global Challenges Research Fund. My main responsibilities are in developing, assessing and communicating intuitive statistical models for predicting prevalence using an array of data sources, including geospatial public health data and chemical analyses.

2019–present
: **Final year project allocation**, *Cardiff University*.
  I have installed a new framework for allocating dissertations to final year students in the School of Biosciences at Cardiff University. The software-based framework has reduced the bulk of their workflow down to a matter of seconds. In doing so, the allocation is both mathematically fair and student-optimal, with almost all students being allocated their first or second choice of project.

2019–2020
: **MMORS dissertation co-supervisor**, *Cardiff University*.
  I assisted in the supervision of a MMORS final year project with Dr Vince Knight conducting an empirical study of Folk Theorems in repeated games. My primary role was to consult on how best to develop the supporting research software, and in the writing process.

2017–2021
: **PhD studentship teaching**, *Cardiff University*.
  Throughout my time as a postgraduate student, I supported a number of modules and services as a tutor and assessment advisor. These roles provided me with countless opportunities to reinforce teaching and learning in a variety of environments, including group, one-on-one, structured and drop-in settings. This range has allowed me to demonstrate my nature as a mathematician and educator, one who is proactive, analytical, logical, and enthusiastic.

✉ *henrydavidwilde@gmail.com* • 🐦 *daffidwilde* • ⌂ *daffidwilde*

Publications and pre-prints

[KPW20]   Vince Knight, Michalis Panayides, and Henry Wilde. *Python for Mathematics*. 2020. DOI: 10.5281/zenodo.4074114.

[Wil+20]   Henry Wilde et al. *Segmentation analysis and the recovery of queuing parameters via the Wasserstein distance: a study of administrative data for patients with chronic obstructive pulmonary disease*. 2020. arXiv: 2008.04295 [stat.AP].

[WKG19]   Henry Wilde, Vincent Knight, and Jonathan Gillard. "Evolutionary dataset optimisation: learning algorithm quality through evolution". In: *Applied Intelligence* 50.4 (2019), pp. 1172–1191. DOI: 10.1007/s10489-019-01592-4.

[WKG20a]   Henry Wilde, Vincent Knight, and Jonathan Gillard. *A novel initialisation based on hospital-resident assignment for the k-modes algorithm*. 2020. arXiv: 2002.02701 [cs.LG].

[WKG20b]   Henry Wilde, Vincent Knight, and Jonathan Gillard. "Matching: A Python library for solving matching games". In: *Journal of Open Source Software* 5.48 (2020), p. 2169. DOI: 10.21105/joss.02169.

Software projects

[Kni+19]   Vince Knight et al. *Blackbook: Black for Jupyter notebooks*. 2019–present. DOI: 10.5281/zenodo.2553362.

[WK18a]   Henry Wilde and Vince Knight. *EDO: A library for generating artificial datasets through evolution*. 2018–present. DOI: 10.5281/zenodo.2552890.

[WK18b]   Henry Wilde and Vince Knight. *Matching: A package for solving matching games*. Software. 2018–present. DOI: 10.5281/zenodo.2553125.

Additional activities

Jan 2020   **SIAM UKIE Annual Meeting 2020**, *The University of Edinburgh*.
Received a travel award to present a poster on evolutionary dataset optimisation — including a case study comparing k-means and DBSCAN clustering.

May 2019   **Welsh Mathematics Colloquium**, *Gregynog Hall*, Powys.
Gave an in-depth talk on the mathematical principles of evolutionary dataset optimisation and some of the issues surrounding algorithm evaluation.

Mar 2019   **Data Science Campus Seminar Series**, *Office for National Statistics*.
Spoke on my work into evolutionary dataset optimisation. In particular, its applications to the field of data simulation and synthesis.

Feb 2019   **NHS Wales Modelling Collaborative**, *South Wales*.
Invited to speak about how my data-driven approach to my research would impact the Cwm Taf Morgannwg Health Board.

2018–2020   **Advanced Python Workshop**, *Cardiff University*.
Founded a group for postgraduate students to engage in monthly, tutorial-based sessions about more advanced aspects of Python (such as parallelisation and automated testing) followed by a code clinic.

2018–2019   **STEMLive**, *Cardiff University*.
Volunteering, designing and running a stall at the annual outreach event for school children in the South Wales region to engage with STEM subjects.

2018   **EURO Support for NATCOR Bursary**.
Financial support to attend NATCOR courses in Approximation Algorithms & Heuristics and Predictive Analysis & Forecasting.

References available upon request

✉ henrydavidwilde@gmail.com   •   🐦 daffidwilde   •   ⬡ daffidwilde