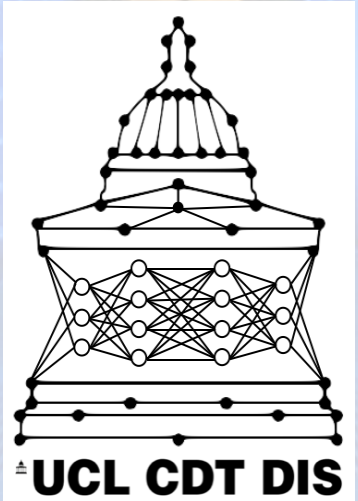
UCL Centre for Doctoral Training (CDT) in Data Intensive Science (DIS)

Jason McEwen

17 December 2020



UCL

Science & Technology Facilities Council

CDT origin & vision

UCL won bid to host STFC's first ever CDT in 2017

1) To produce graduates with advanced and widely applicable DIS skills
   - Ready for a dual career, either in academia or in industry

2) To become a platform that
   a) fosters collaboration across DIS disciplines within UCL and beyond, to accelerate the development and application of pioneering DIS techniques
   b) aids knowledge exchange and cross-fertilization of novel DIS ideas and technologies between industry and academia

3) Enrich the STFC science programme through the development and/or application of advanced DIS tools and expertise

Fast forward 3 years...

- We are about to recruitment of 5th student cohort
  - Currently we have 16+9+9+11=**45** PhD students in the programme
- STFC funding for 24 studentships (~£3M)
  - And another 21 studentships co-funded by UCL and industry partners
- Currently ~80 academics actively engaged in the Centre
- More than 20 partners actively engaged in the Centre
  - Spanning a broad range of industries, as well as academic/research partners
- Newton Fund Capacity Building in the Middle East initiative (£375k)

Unique features of the CDT programme

A 4-year PhD that includes a 6-month placement with one of our partners, working on DIS/Applied AI projects other than the students' PhD research.

- Cohort-based training, front-loaded in 1st year with in-depth courses in Scientific Computing, Statistics, Machine Learning and other DIS techniques
- Industry-led Group Projects (data/challenges proposed by partners)
- Co-supervision with academics in Comp. Science, Statistics, Elec. Engineering
- Research in some of the world's most complex and data intensive experiments in Particle Physics and Astronomy

The halo effect – added benefits

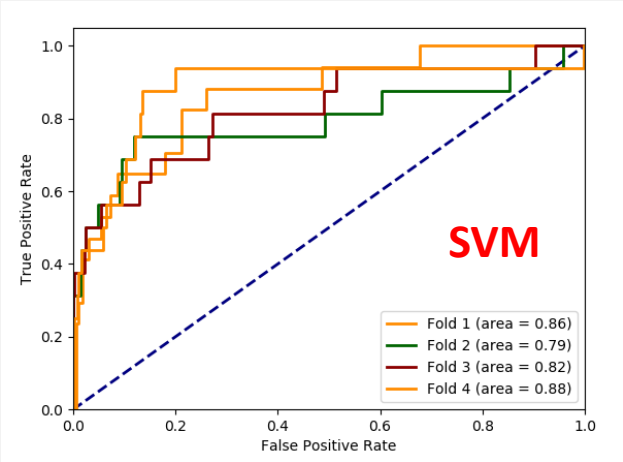
- Training activities, as well as CDT events open to non-CDT PhD students and RA's (networking/career opportunities)

- Introduced the scheme of Innovation mini-Fellowships for finishing non-CDT students and RA's
  - 3-6 months of placements with our industry partners, for those who want to explore the option of a career in industry

- In summer 2018 we organized the 1st STFC national school in "Artificial Intelligence for Data Intensive Science and Industry"
  - 120 participants; most training provided by Industry partners

Example industry group project - TfL

- Aim: to predict train failures 48h before they would happen
- Data: 28M time-stamped “events” from TfL’s Victoria line trains
  - Each containing 717 sensor readings
- Four 1st year students worked for 3 months at 30% of their time
  - Cleaned and preprocessed the data
  - Tested large number of ML algorithms
    - BDTs, ANNs, SVMs
- Outcome: achieved ~30% True positives for ~1% false positives using SVMs
  - TfL were impressed!



CDT management team

Centre Co-Directors: Prof. Nikos Konstantinidis & Prof. Ofer Lahav


Directors of Research: Prof. Jason McEwen & Dr Tim Scanlon


Directors of Training: Prof. Jonathon Tennyson FRS & Prof. Amelie Saintonge


Admissions & Graduate Tutor: Dr Ingo Waldmann & Dr Anasuya Aruliah


CDT partners

ESA-UCL co-sponsored studentship

- First ESA and UCL CDT-DIS shared PhD studentship focusing on ML development in large, unlabelled data sets.
- Project will focus on unsupervised and semi-supervised patter recognition of large ESA catalogue data (e.g. Hubble, GAIA).
- Student: David Smith. Supervisors: Bruno Merin (ESA/ESAC), Ingo Waldmann (UCL)
- A very active collaboration with the ML group at ESAC and UCL, building on existing work with ESA data at UCL.
- A large number of CDT students now developing ML solutions using ESA data (e.g. Gordon Yip, Mario Morvan, Max Hipperson, Arianna Saba, David Smith)
- Great foundation for further ESA-UCL collaboration in AI.