Applied machine learning and data mining

Contact:

Dr Neil Davey, n.davey@herts.ac.uk,
Dr Yi Sun, y.2.sun@herts.ac.uk

Projects and Requirements

We are currently involved in six collaborative projects, investigating how recently developed machine learning techniques can help in the analysis of real data. The areas are:

1. Developing computational models for astronomical sources with variable brightness. Such variability takes many forms, e.g small dark spots on rotating stars like the sun, pulsating stars, eclipsing binary stars, proto-stars with variable luminosity from infalling matter, nova and supernova explosions, quasars powered by matter sucked into the central black hole. The goals are 1) classifying variable astronomical sources by light curve shape, brightness and colour and 2) measuring the period or any other significant timescales. This is a collaboration with the School of Physics Astronomy Maths at the University of Hertfordshire. The astronomy group is co-leading the first major infrared study of the variable sky, known as VVV, and they have interests in the LSST project, a major new facility for the 2020s. A couple of descriptions of background for this project can be found at: https://www.epj-conferences.org/articles/epjconf/abs/2017/21/epjconf_puls2017_03001/epjconf_puls2017_03001.html and http://www.astroscu.unam.mx/rmaa/RMxAC..49/PDF/RMxAC..49_oral47.pdf.

2. In the past 20 years LIDAR (LIght Detection And Ranging) has become an important remote sensing technology for studying properties of air pollution and its impact on climate and weather. Optical data from LIDAR measurements as well as simulated data will be used to investigate if properties of particle pollution, such as mean particle size and complex refractive index, can be inferred. Such parameters are a crucial piece of information needed in studying the impact of air pollution on climate change. Another task will be to investigate the information content of optical data collected by LIDAR which will help optimizing this measurement technology, for example for airborne and space-borne applications. This is a collaboration with the School of Physics Astronomy Maths at the University of Hertfordshire.

3. Developing computational models for characterizing small particles based on their 2-dimensional light scattering patterns. This is a collaboration with the School of Physics Astronomy Maths at the University of Hertfordshire. A couple of descriptions on background of this project can be found at: http://www.atmos-chem-phys.net/14/1649/2014/acp-14-1649-2014.html and http://www.sciencedirect.com/science/article/pii/S0022407317302443.

4. Predicting the rate at which drugs and other chemicals are absorbed by human skin. This is a collaboration with the School of Pharmacy at Keele University. A description of our work can be found at: http://www.tandfonline.com/doi/abs/10.1080/1062936X.2015.1018941.

5. Classifying signals carried along optical fibre in order to correct distortion. This is a collaboration with the Institute of Computational Technologies, at Novosibirsk, Russia. A description of this project, entitled ‘Reducing bit error rate of optical data transmission with neighbouring symbol information using a linear support vector machine’, can be found at: https://aaltodoc.aalto.fi/bitstream/handle/123456789/18224/isbn9789526064437.pdf.

6. Automatically segmenting objects in astronomical images. This is a collaboration with the Centre for Astrophysics Research at the University of Hertfordshire. A recent publication on

this work, entitled ‘An automatic taxonomy of galaxy morphology using unsupervised machine learning’, can be found at: https://academic.oup.com/mnras/article/473/1/1108/4159372.

In these projects we are using a variety of techniques, such as: Deep Neural Networks, Gaussian Processes and Support Vector Machines. We are looking for a PhD student who would be interested in working on one of these projects. The prospective candidates should have a strong background in Computer Science or another relevant computational discipline. In particular, they should demonstrate strong programming skills in one or more major computer languages. Ideally such a student would have experience of both machine learning and have some knowledge of the application domain. However we will consider applications from candidates with either a background in machine learning or knowledge of the problem domain.