ACCELERATE PROGRAMME FOR SCIENTIFIC DISCOVERY

Network Event
Wednesday 29 March 2023

Flash Talk Abstracts

Automatic Detection for Bioacoustics

Arik Kershenbaum, Official Fellow and College Lecturer, Girton College, University of Cambridge, and Department of Zoology

Endangered species often live in environments where traditional monitoring is difficult and requires invasive capturing and sedating the animal to attach GPS collars. Passive acoustic multilateration gives the precise location of each vocalising individual by comparing the time of arrival of sound at multiple recording devices at different locations. However, identifying sounds within the recording is currently performed manually, and this is clearly completely unscalable for a realistic deployment. It is essential to automate the scanning of recordings and the identification of vocalisations in near real-time. Such a development would be revolutionary for biodiversity conservation, translatable also to other species and systems. Several research groups have used deep neural networks and other artificial intelligence techniques to detect and classify animal calls automatically. Some progress has been made with high-quality recordings at close range, but major new developments are needed to leverage the possibilities of passive localisation under natural field conditions. This project will convene a 4-day workshop bringing together biologists and computer scientists already working in this field. We will generate new impetus and new collaborations to pool effort and allow common problems to be addressed by a multidisciplinary collaboration, rather than in isolation.

What can causality offer the environmental sciences?

Sebastian Hickman, PhD student, AI for the study of Environmental Risks CDT and Yusuf Hamied Department of Chemistry

Distinguishing correlation and causation is a fundamental problem in many scientific fields. Particularly when experiments are unethical or infeasible, we often rely on observational data to infer causal relationships. The environmental and climate sciences are prime examples of fields where direct experimentation is difficult, and therefore may benefit from techniques to infer causality from observational data. However, causal methods are not yet widely used in these fields, and we have therefore been building a community of both data scientists and environmental scientists to highlight the possible benefits of using causal approaches, supported by the Accelerate Programme. This flash talk will provide some examples from environmental and climate science where correlation and causation need to be disentangled, alongside an overview of how these approaches may help researchers to identify robust causal relationships from data, and better inform policy-makers.

A new eye on semantics. Coregistration of eye-movements and MEG/EEG to study semantics in context

Federica Magnabosco, PhD student, MRC Cognition and Brain Sciences Unit

I will talk about my PhD in cognitive neuroscience, specifically in the field of semantic cognition, which is the study of how our knowledge is represented in the brain and how we can use it according to context and task constraint. Focusing on the brain systems that support language and semantic memory, I'm interested in how sentence context influences the way we process single-word meaning. After an overview of the project, I will focus on those parts of my work where what I learned at the Data Science Residency and the Machine Learning Academy were most helpful, including the creation of the stimuli, synchronization, and pre-processing of the data.

ML-powered brain-computer interfaces in clinical applications

João Araújo, PhD Student, Department of Psychology

In this flash talk, I will present a brief introduction to the work I develop with non-invasive brain-computer interfaces (BCI) for clinical applications – namely their potential application in the remediation of developmental language disorders. I will outline how data science and machine learning are central in transforming brain oscillations into a clinically-relevant controller. Finally, I will share how the accelerate science programme helped me not only creating and open-sourcing relevant BCI projects with deep learning but also getting a podium on a worldwide Kaggle competition.

Understanding biology in the age of artificial intelligence

Matthew Greenig, PhD Student, Yusuf Hamied Department of Chemistry

Modern biological research is becoming increasingly reliant on machine learning (ML) approaches, particularly for deriving insights into the behaviour of systems and processes that cannot be tractably understood with traditional scientific techniques. ML methodologies are useful for modelling complex and non-linear relationships – ubiquitous in biology – but typically produce models that are not straightforward to understand or interrogate. While the opacity of ML models is not an issue in many domains, their emerging role as tools in fundamental biological research poses important questions regarding their place in the traditional picture of scientific understanding. As a collaboration between researchers in ML theory, biological research, and epistemology, our work explores the consequences and implications of biological scientists relying on ML models in their research. We tie together themes in theoretical ML research with philosophical conceptions of knowledge and understanding, and summarise how those ideas relate to the application of ML in research areas including protein folding and single cell genomics. We hope that our ideas can guide future applications of ML in biology to produce research that is impactful, reproducible, and reliable.

18:30 – 19:30 Poster and Networking Session

1 Uncovering the THz molecular component of surface enhanced Raman spectra

Alexandra Boehmke, PhD student, Department of Physics

Using machine-learning clustering, surface-enhanced Raman spectra in the terahertz region (0.1 – 6 THz) are shown to be molecule specific. Surface-enhanced Raman spectroscopy (SERS) uses a metal nanostructure to funnel light into a sub-diffraction-limited volume. This enables Raman spectroscopy of a small number of molecules adsorbed onto the metal surface, from ~100 down to a single molecule. Whereas the standard range of frequencies observed, the infrared region of the spectrum, characterizes the component functional groups of a molecule, the THz region of the Raman spectrum provides information about molecular structure and environment. Combining techniques for THz Raman and surface enhancement opens up the potential to observe short-range order in small sample volumes, such as protein conformation, and structural changes for applications such as molecular electronics. A significant obstacle to the use of THz SERS is the extraction of the molecular signal from the measured spectrum, which also contains components originating from the metal nanostructure. Based on a physical model of electronic Raman scattering, we fit a Bose-Einstein distribution as a background to the molecular spectrum. We show that the ERS-removed spectrum is molecule dependent by clustering the THz region of the spectra and finding perfect correlation with the true molecular labels.

2 Discovery of Potent Inhibitors of α-Synuclein Aggregation Using Structure-Based Iterative Learning

Robert Horne, PhD student, Yusuf Hamied Department of Chemistry

Machine learning methods hold the promise to reduce the costs and the failure rates of conventional drug discovery pipelines. This issue is especially pressing for neurodegenerative diseases, where the development of disease-modifying drugs has been particularly challenging. To address this problem, we describe here a machine learning approach to identify small molecule inhibitors of α-synuclein aggregation, a process implicated in Parkinson's disease. Because the proliferation of α-synuclein aggregates takes place through autocatalytic secondary nucleation, we aim to identify compounds that bind the catalytic sites on the surface of the aggregates. To achieve this goal, we use structure-based machine learning in an iterative manner to first identify and then progressively optimize secondary nucleation inhibitors. Our results demonstrate that this approach leads to the facile identification of compounds two orders of magnitude more potent than existing ones.

3 Dimensionality Reduction: A Probabilistic Perspective

Aditya Ravuri, PhD student, Accelerate Programme, Department of Computer Science and Technology

We present a probabilistic framework that gives rise to many dimensionality reduction algorithms as inference processes of our framework.

4 PKSmart: A Public Tool to Predict In Vivo Pharmacokinetics of Small Molecules

Srijit Seal, PhD student, Yusuf Hamied Department of Chemistry

It is not only the biological activity of a compound that determines its mechanism of action but also the exposure. Drug exposure can be defined using human pharmacokinetics (PK) parameters that affect the blood concentration profile of a drug, such as steady-state volume of distribution (VDss), total body clearance (CL), half-life (t½), fraction unbound in plasma (fu) and mean residence time (MRT). Early assessment of PK properties is crucial, such as in DMTA cycles, and this is what we enable in this work with models needing only chemical structure as input. We developed a web-hosted application PKSmart (pk-predictor.serve.scilifelab.se) which users can access using a web browser with all code also downloadable for local use. To the best of our knowledge, this is the first work that releases PK models publicly that can predict human and animal PK parameters using inputs of chemical structure alone.

5 Learning the geometry of complex manifolds

Justin Tan, PhD student, Accelerate Programme, Department of Computer Science and Technology

The metric tensor of a manifold describes the size and shape of the space. A special class of these spaces, Calabi-Yau manifolds, are of particular importance in string theory and mathematics. Finding metrics with the property of Ricci-flatness on a Calabi-Yau is a long standing problem in geometry with deep implications for string theory and phenomenology. One approach to this problem uses neural networks to engineer differentiable approximations to the Calabi-Yau metric. From this geometrical/topological data may be computed to derive phenomenological parameters associated with the low-energy reduction of a string theory, bridging string theory with observable physics.