Tackling environmental change using data science and AI: Report of the Environmental Intelligence Summit, May 2019

Our changing environment presents a series of inter-related challenges that will affect everyone's future health, safety and prosperity. Environmental Intelligence is the integration of environmental and sustainability research with data science, artificial intelligence and cutting-edge digital technologies to provide the meaningful insight that is required for individuals, businesses and policy-makers to address these challenges and mitigate the effects of environmental change.

To boost the integration of Data Science and Artificial intelligence (AI) in environmental research and decision-making the University of Exeter hosted the inaugural Environmental Intelligence Summit on 22 May 2019. The event brought together many practitioners and experts for the first time.

A key message of the summit was that Data Science and AI are crucial for our ability to understand the environmental challenges we face, and for identifying workable solutions. It is important that scientists, stakeholders and funding agencies work together to maximise the potential of this new field.

"AI and machine learning are helping us understand and deal with the challenges of climate change," said Murray Simpson, climate risk and resilience lead at IBM.

Following the Summit an Environmental Intelligence Network was established that will allow practitioners and stakeholders to work together and share knowledge on an international scale.

The University of Cambridge and the University of the University of Exeter have recently received funding to establish two Centres of Doctoral Training (CDT) in this area: the UKRI CDT for the study of Environmental Risks (AI4ER) at Cambridge and the UKRI CDT in Environmental Intelligence at Exeter. Both have welcomed their first students in September 2019 and are recruiting students for 2020.

The problem: environmental complexity

Many environmental problems are particularly challenging due to the inherent complexity of the underlying earth systems, such as the climate and biosphere.

UNIVERSITY OF EXETER | INSTITUTE FOR DATA SCIENCE AND ARTIFICIAL INTELLIGENCE

The Alan Turing Institute

2019IDSAI001

"We have more datasets available," said Yaera Chung of the United Nations Development Programme. The trouble is making sense of it all, as traditional forms of data analysis often struggle to identify the most salient patterns.

Examples of highly complex and multi-faceted environmental problems include:

- Ensuring energy security while cutting greenhouse gas emissions to zero. Grid operators must use multiple energy sources, many of them unpredictable.
- Averting biodiversity loss. There are many measures of biodiversity, from the population of a species to the richness of a habitat, and boosting one does not necessarily improve the others.
- Identifying illegal activities such as illegal fishing that cause environmental harm, but which are difficult to track due to remoteness.
- Obtaining and harnessing local-scale data on environmental threats such as air pollution.
- Producing useable information about risk from more abstract data: for instance, inferring flooding risk from information on future rainfall, river catchments and building practices.

The solution: data science and AI

The human brain struggles to integrate complex and multifarious sources of information. Raia Hadsell of Google Deepmind compared truly intricate problems to "a joystick with 27 knobs": something no human could consciously control. However, computers do not suffer this limitation, so can be used to help us understand the intricacies of the Earth system: an endeavour NERC has dubbed the "digital environment".

AI can harness the power of the most advanced computers to understand complex systems. In some cases they can far surpass the abilities of even the most skilled humans. For example, Google DeepMind's program AlphaZero achieved superhuman skill in the board games chess, shogi and Go within 24 hours.

There are many forms of AI, but they can be broadly divided into two categories: those that learn purely by experience, and those that come pre-loaded with information about the system they are studying. For instance, models of the climate system typically include detailed simulations of known physical processes such as evaporation, and phenomena like rainfall emerge from these lower-level processes. In contrast, programs like AlphaZero are not hard-coded with information about the games they play, but learn solely by repeatedly playing the game. This second approach can be useful when we do not clearly understand the processes that are driving a system.

To date, AI has not been widely used in environmental science but with the enormous volumes of data now being collected, and the growing recognition of the complexity of the challenges, the value of these technologies is starting to be recognised.

"Not only are the types of datasets we're using changing, the existing datasets are growing and changing in terms of density," said Kari Dempsey, head of data science at the UK Hydrographic Office. "The density is hugely increasing. For people to review that data in a useful way is becoming hugely challenging."

Speakers at the Summit agreed that the wide-scale integration of data science and AI within environmental science has the potential to drive significant progress across a wide variety of environmentally-related problems.

2019IDSAI001

Opportunities for Environmental Intelligence

During discussions at the Environmental Intelligence Summit, experts in the field identified a number of significant opportunities for innovative data science using artificial intelligence. These included:

- More joined-up thinking about environmental problems. In particular, AI could help identify novel pathways for society that preserve multiple Earth systems, instead of focusing solely on a single system like climate or water. For example, the RCP2.6 scenario is a standard tool for climate modellers and is designed to limit the global temperature rise to well below 2 °C – but it achieves this through an enormous expansion of bioenergy, which requires huge areas of land and would therefore have major impacts on biodiversity.
- Simulations could be used to estimate the scope for future extremes, for instance the potential scale of future rainfall events influenced by climate change. This would aid risk planning. Estimating how severe an extreme event could be is currently challenging, as our records are relatively short. "The January 2014 Thames flooding was way outside the observed pack,…" said Julia Slingo, former chief scientist of the UK Met Office. However, a 2017 study she co-authored showed that it could have been anticipated: due to our greenhouse gas emissions, "…in south east England there is a 7% chance of exceeding the current rainfall record in at least one month in any given winter".
- Personalising environmental issues to bring their significance home to people. For instance, simulations of airflow in a city could be used to supply individual households with personalised information about air quality.
- Combining physics-based models, used in fields where we have good understanding of the underlying mechanisms that are driving environmental processes, with machine learning for fields where we have data but limited understanding. This would improve our ability to make forecasts under uncertainty.
- Identification of critical biodiversity hotspots by analysing the increasing volume of data from environmental DNA.
- Improving the design of key infrastructure such as smart cities, energy grids and sustainable agriculture.

However, delegates also emphasised that, before Environmental Intelligence can flourish as a field of study and practice, a number of challenges must be overcome.

Challenges for Environmental Intelligence

Many of the biggest barriers for the use of data science and AI are factors outside of the scope of the computer programs. For instance, in many environmental fields, data is extremely heterogeneous and data sets often need a lot of cleaning up before they can be meaningfully compared. Similarly, the things policy aims to maximise – like ecological integrity and human wellbeing – are often hard to measure, particularly compared to traditional measures of success like Gross Domestic Product (GDP).

However, experts at the Summit agreed that the real challenges are political and ethical. A machine-learning tool like AlphaZero can only learn when it is given a target, known as an objective function. In a game of Go, the objective function is easy to define. However, if such a program is used to design an environmental policy, the all-important choice of objective function will be fundamentally subjective and value-based. In other words, while the workings of the program may not be political, the outcome it aims for unquestionably is. It will be crucial to set such objective functions in a democratic and inclusive way, so as to reduce injustice rather than inadvertently entrenching it. Similar questions arise around who should control the underlying data.

Finally, there is also the question of the environmental impact of AI; data centres consume a large quantity of energy, which must be minimised and ultimately decarbonised to avoid contributing to climate change. Hadsell described research by DeepMind in which an AI algorithm was given control of a Google data centre and reduced energy consumption for cooling by 40%.

To ensure that artificial intelligence is used as skilfully and ethically as possible, a broad coalition of stakeholders needs to be involved. Discussions at the Summit suggested that these groups should include, at minimum:

- Scientists
- Businesspersons
- Policymakers and lawyers
- Diverse civil society groups, social forums and citizens' assemblies
- Educators
- Economists and social scientists
- The media

However, one delegate suggested that the true list of stakeholders is, simply, "everyone".

Authors: Michael Marshall (www.michaelcmarshall.com)
Prof Gavin Shaddick, University of Exeter