

Next generation of statistical tools to be developed for the big data age

September 21 2016

Statisticians are developing new ways to interpret the unprecedented amounts of data being generated continuously all around us.

Whether it is smart phones packed full of sensors measuring your health, location and the weather, or connected cars driving past roadside monitors measuring traffic volumes or air quality, or even thousands of sensors on individual oil wells, in the world of the 'Internet of Things' and 'big [data](#)' massive streams of information are now being generated and collected at an unprecedented scale.

It is estimated that by 2020 there will be more than 30 billion devices collecting data streams. Being able to interpret and take advantage of all this data will lead to great economic and societal benefits – providing advances in areas such as e-health and communications and enabling more of us to lead healthier and more productive lives. A CEBR report has estimated that big data will be worth up to £40 billion to the UK economy by 2017.

This new form of data brings with it fundamentally new data analytic challenges. For example, while traditional statistical methods were suitable, and readily computed, for modest amounts of data, they were not developed with the streaming data age in mind. To address this, a new programme of research is being funded by the Engineering and Physical Sciences Research Council. The £2.75M 'StatScale: Statistical Scalability for Streaming Data Pathways For Impact' programme is being led by members of Lancaster University's Data Science Institute in

partnership with colleagues in the University of Cambridge's Statistical Laboratory.

Idris Eckley, Professor of Statistics at Lancaster University, said: "The ubiquity of sensors in everyday systems and devices, such as smart watches to instrumented oil fields, means there is enormous potential for societal and economic benefit if information can be extracted effectively.

"The volume, scale and structure of this contemporary data poses fundamentally new and exciting statistical challenges that cannot be tackled with traditional methods. Our aim is to develop a paradigm-shift in statistics, providing a new statistical toolbox to tackle, and capitalise on, these huge data streams."

Professor Richard Samworth from the University of Cambridge said: "Many classical methods are either impractical or not fit for purpose for dealing with these [data streams](#). StatScale will develop the theoretical and methodological foundations that will underpin the next generation of scalable statistical algorithms. These methods are urgently required if the UK is to maintain its competitive edge across a range of scientific and industrial challenges."

StatScale benefits from significant partnerships with industry. Companies including Shell UK, BT, AstraZeneca and the Office for National Statistics have agreed to trial new methods and models that emerge from the programme so that they can be rapidly tested and refined in real-world situations. These collaborations will both help inform StatScale's research agenda and will also help this research make direct economic and societal benefit swiftly.

Professor Tom Rodden, Deputy Chief Executive of The Engineering and Physical Sciences Research Council (EPSRC) said; "Every day,

individually and collectively, we are generating and contributing to vast quantities of information, this is the 'big data' age. However, to make effective use of this data, that will bring economic and [societal benefits](#), we must have reliable, accurate methods of interpreting it.

"The StatScale project, which benefits from close collaborations with industrial partners, will be key to providing the statistical tools needed to harness this information revolution."

StatScale will be led by Professors Idris Eckley and Paul Fearnhead from Lancaster University, and Professors John Aston and Richard Samworth of the University of Cambridge.

Provided by Lancaster University

Citation: Next generation of statistical tools to be developed for the big data age (2016, September 21) retrieved 23 May 2024 from <https://phys.org/news/2016-09-statistical-tools-big-age.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.