Prospects in Data Science.
A multidisciplinary symposium

Tuesday 12-14 January 2016

Confirmed speakers

John Ashton	Peter Grindrod	Sofia Olhede
Gunnar Carlsson	Yike Guo	Nigel Shadbolt
Fai Cheng	Wendy Hall	Anne Trefethen
Dave Coplin	David Hand	Frank Woods
Mark Girolami	Vitalyi Kurlin	
Arthur Gretton	Jane Naylor	

Carey's Manor Hotel, Brockenhurst, SO42 7RH

Please book on eventbrite

datascience.soton.ac.uk/conference2016

Welcome to the Data Science Symposium 2016

Introduction

In the current Information Age, data has become a commodity that is driving development crucial to future economic success, particularly for service-based economies such as the UK. The potential to transform the economic landscape is tantalising, from providing business with strategic advantage or new services, to revolutionising medical diagnostics, among many other benefits to society. However this potential cannot be realised unless new methods for handling, analysing, and extracting knowledge from data are made available. This is particularly relevant in the context of Big Data, where scalable techniques and algorithms are vitally important. The emerging field of Data Science usually refers to the interface between Statistics, Mathematics, and Computer Science that is providing the much sought novel techniques and approaches arising from the cross-fertilisation of ideas between these complementary domains. Data Science is rapidly gathering momentum, and suggests promising new research avenues in the near future. In recognition of this momentum, EPSRC have established the Alan Turing Institute to promote advanced research and translational work in the application of data science, acknowledging that this requires leadership both in advanced mathematics and in computing science.

Set in the heart of the gorgeous New Forest, this Data Science Symposium organised by the University of Southampton brings together a multi-institutional, high-profile panel of speakers to promote the cross-fertilisation of ideas between the different domains of Data Science and discuss the prospects of this emerging field in the near future.

This event is financed through the EPSRC Institutional Sponsorship grant 'Southampton Data Science'.

For more information about Data Science at the University of Southampton please visit www.southampton.ac.uk/datascience

Organisers

Main organiser: Ruben Sanchez-Garcia (Mathematical Sciences)
'Southampton Data Science' Co-ordinator: Tricia Worby
Grant co-applicants and other organisers:

Southampton Statistical Sciences Research Institute	Jon Forster, Steve Gilmour
Mathematical Sciences	Jacek Brodzki, Joerg Fliege, Jelena Grbic, Rebecca Hoyle, Ben MacArthur, Graham Niblo, Chris Potts, Kostas Zygalakis
Electronics and Computer Science	Mahesan Niranjan, Chris Phethean, Elena Simperl, Thanassis Tiropanis
Institute for Life Sciences	Peter J Smith
Administrate Data Research Centre for England	Peter W Smith
Centre for Biological Sciences	Rob Ewing, Paul Skipp
Chemistry	Jeremy Frey
Medicine	Ratko Djukanovic
Geography and Environment	Andrew Tatem

Data Science Symposium Programme

Day 1 (Tuesday 12th January)

Time	Speaker	Affiliation	Title
10.00 – 10.30	**Registration/Coffee**		
10.30 – 11.30	David Hand	Imperial College London	The roles of models in data science
11.30 – 12.30	Peter Grindrod	University of Oxford	Red Herrings and Wild Goose Chases - Creating Analytics for Impact
12.30 – 1.30	Lunch		
13.30 – 14.10	Wendy Hall	University of Southampton	Observatories and data analytics for Web Science
14.10 – 14.50	Frank Wood	University of Oxford	Probabilistic Programming
14.50 – 15.30	John Aston	University of Cambridge	When Big Data isn't "big enough": tales from an infinite data world
15.30 – 16.00	Tea/Coffee		
16.00 - 17.00	Fai Cheng	Lloyd's Register	Big Data and its Transformational Effects
17:30 –18:30	Drinks Reception		

Day 2 (Wednesday 13th January)

Time	Speaker	Affiliation	Title
9.00 – 9.30	**Coffee**		
9.30 – 10.30	Gunnar Carlsson	Stanford University	Topology for Data Analysis
10.30 – 11.00	Tea/Coffee		
11.00 – 11.40	Vitaly Kurlin	Durham University	Topological Computer Vision
11.40 – 12.20	Jane Naylor	Office of National Statistics	Data science and its role within official statistics
12.20 – 14.00	Lunch		
14.00 – 15.00	Anne Trefethen	University of Oxford	Turning data science theory into data science practice
15.00 -15.40	Sofia Olhede	University College London	Big Data in Time
15.40 – 16.10	Tea/Coffee		
16.10 – 17.00	Poster Storm		
17.00 – 18.00	Poster Session		
18.30 – late	Conference Dinner		

Day 3 (Thursday 14th January)

Time	Speaker	Affiliation	Title
9.00 – 9.30	**Coffee**		
9.30 – 10.30	Arthur Gretton	University College London	Kernel nonparametric tests of homogeneity, independence and multi-variable interaction
10.30 – 11.00	Tea/Coffee		
11.00 – 12.00	Yike Guo	Imperial College	Big Data for Better Science
12.00 - 1.30	Lunch and End		

Speakers and Abstracts

David Hand
The roles of models in data science
Abstract:

There have been claims that the new world of big data means that models are no longer necessary – we can simply look at the data and see how the world is, without constructing doubtful mechanisms. However, this oversimplifies things. In particular, it fails to note the distinction between substantive and empirical models. The 'big data revolution' has been largely driven by empirical models, but they are not the answer to all our problems.

Peter Grindrod
Red Herrings and Wild Goose Chases - Creating Analytics for Impact
Abstract:

We will discuss the highly distinctive nature (and potential) of the knowledge that is derived via the application of analytics processes to big data resources (that themselves are almost always repurposed). How is this district from science? We shall consider Pragmatics versus Theory: what works and why should anything work? This includes discussing the proper role for academic research versus that of commercial players' own R&D (i.e. that universities and institutes don't compete with commercial opportunities and ventures). We shall make a clear distinction between various data science sectors: from highly regulated areas (where rigorous requirements and methods has increased costs and thus broken the business model) to the "wild west" (where public participation, ethics and public perception require a clear and well understood balance of benefits). We will consider Big Science versus Big Economic Impact - why do we confuse these? Where is the confusion? What should a national strategy for data science contain? We shall discuss some possible fields for analytics applications, focusing on those sectors that can build novel services, products and jobs: *what does success for the UK look like*?

Wendy Hall
Observatories and data analytics for Web Science
Abstract:

Over the last 25 years the Web has evolved into a critical global infrastructure. Since its emergence in the 1990s, it has exploded into hundreds of billions of pages that touch almost all aspects of modern life. Little appreciated, however, is the fact that the Web is more than the sum of its pages and it is more than its technical protocols. Vast emergent properties have arisen that are transforming society. Web Science is the study of the Web as a socio-technical system. As the Web becomes increasingly significant in all our lives, studying it from an interdisciplinary perspective becomes even more important.

We are now rapidly moving into a world of data on and about the Web, which gives rise to even more opportunities and challenges. In this talk, we will explore the role of Web Science in helping us understand the origins of the Web, appreciate its current state and anticipate possible futures in order to address the critical questions that will determine how the Web evolves as a social-technical network. We will discuss the role of observatories and data analytics for the development of new methodologies for longitudinal research in Web Science.

Frank Wood
Probabilistic Programming
Abstract:

Probabilistic programming is a general-purpose means of expressing and automatically performing model-based inference. A key characteristic of many probabilistic programming systems is that models can be compactly expressed in terms of executable generative procedures, rather than in declarative mathematical notation. For this reason, along with automated or programmable inference, probabilistic

programming has the potential to increase the number of people who can build and understand their own models. It also could make the development and testing of new general-purpose inference algorithms more efficient, and could accelerate the exploration and development of new models for application-specific use.

John Aston

When Big Data isn't "big enough": tales from an infinite data world

Abstract:

Functional Data Analysis (FDA), or the statistical analysis of continuous curves and surfaces, has emerged over the last 20-30 years as not only an area of statistics with interesting theory but also an abundance of exciting applications. As an applied statistician, using the concepts for FDA analysis allows various scientific fields to be investigated in a similar way. In this talk, I will take a tour of some of the areas where using FDA has proved useful in my work. These areas range from understanding the human brain, to exploring the evolution of languages. It will be seen that while the applications are very different, a similar FDA point of view arises naturally in each case.

Fai Cheng

Big Data and its Transformational Effects

Big Data are now woven into our personal lives and society. Big Data and its potential to create value needs to be taken seriously by every industry sector in the global economy if they want to compete, and, like other essential factors of production such as lands, labour, capital, much of modern economic activity simply could not take place without them. The use of Big Data — large pools of data that can be brought together and analysed to discern patterns and make better decisions — will become the basis of competition and growth for individual firms, enhancing productivity and creating significant value for the world economy by reducing waste and increasing the quality of products and services. For countries, Big Data is the "new oil" on which wealth and job creation will depend. For individuals and organisations generating the Big Data, they need to be aware of their use and potential impact on them too.

Gunnar Carlsson

Topology for Data Analysis

Abstract:

The study and analysis of large and complex data sets is one of the most challenging intellectual problems of our time. There are many mathematical approaches to it. In this talk, we will discuss a methodology based on the mathematical study of generalized shape, known as topology. The methods provide a quick and simple way to get information about the data set, as well as a collection of modeling tools. We will present the ideas, with numerous examples, many from the life sciences.

Vitalyi Kurlin

Topological Computer Vision

Abstract:

Topological Computer Vision is a new research area within Topological Data Analysis on the interface between algebraic topology and computational geometry. The flagship method of persistent homology quantifies topological structures hidden in unorganized data across all scales. The talk will review recent applications to Computer Vision including auto-completion of contours, parameterless skeletonization and superpixel segmentation of images. The last work is joint with Microsoft Research Cambridge and was funded by the EPSRC Impact Acceleration Account through a Knowledge Transfer Secondment.

Jane Naylor
Data science and its role within official statistics
Abstract:
The Office for National Statistics (ONS) is the UK's largest independent producer of official statistics. It is responsible for collecting and publishing statistics related to the economy, population and society at national, regional and local levels. These statistics are used within central and local government to underpin policy making and to plan services. The ONS recognises the importance of examining the potential of data science. This presentation will cover progress being made at the ONS to investigate the opportunities as well as the challenges of using big data sources and associated methods and technologies within official statistics. The future impact of data science for official statistics will also be discussed. Particular focus will be given to case studies from ONS such as
- Web scraping data for price statistics
- Using geo-located Twitter data to provide intelligence on migration

Anne Trefethen
Turning data science theory into data science practice
Abstract:
Being the Chief Information Officer for the University of Oxford I have the opportunity to take some of the theories and lessons from data science research applications to practical, real-life problems. I will discuss some of the opportunities and challenges this brings through exemplars including cybersecurity, and the implications for data science infrastructure at scale.

Sofia Olhede
Big Data in Time
Abstract:
Climate observations often take the form of multiple time series observed at many spatial locations. My group has been studying such data in the form of measurements of global circulation using drifting instruments, or satellite-tracked ocean surface drifters. I will discuss what new models and inference methods are required to understand such observations, and the computational challenges we have resolved.

Arthur Gretton
Kernel nonparametric tests of homogeneity, independence and multi-variable interaction
Abstract:
I consider three nonparametric hypothesis testing problems: (1) Given samples from distributions p and q, a homogeneity test determines whether both distributions are the same (like a t-test but more general); (2) Given a joint distribution p_{xy} over random variables x and y, an independence test investigates whether x and y are dependent (like a Pearson correlation test but more general); (3) Given a joint distribution over several variables, we may test for whether there exists more complex interactions, for example where two independent causes individually have weak influence on a third dependent variable, but their combined effect has a strong influence. I will present nonparametric tests for the three cases described, based on distances between embeddings of probability measures to reproducing kernel Hilbert spaces (RKHS). The tests benefit from decades of machine research on kernels for various domains, and thus apply to distributions on high dimensional vectors, images, strings, graphs, groups, and semigroups, among others. Finally, time permitting, I will briefly discuss supervised learning with (samples from) probability distributions as inputs.

Dave Coplin
The Rise of the Humans: How to Outsmart the Digital Deluge
Abstract:
In this session we will explore how the rapidly rising volume of information is affecting all aspects of our lives as individuals, as consumers and customers, as workers, and in business. We'll look at the problems we face, and the mistakes and assumptions we make and how we should be using technology to reimagine how we live, work and do business.

Yike Guo
Big Data for Better Science

First name	Second name	Affiliation	Email
Reham	Al Tamime	University of Southampton	rat1g15@soton.ac.uk
Gregory	Ashton	University of Southampton	ga7g08@soton.ac.uk
John	Aston	University of Cambridge	jada2@cam.ac.uk
Alistair	Bailey	University of Southampton	ab604@soton.ac.uk
David	Barber	DSTL	dbarber@dstl.gov.uk
Mariano	Beguerisse	Oxford University	beguerisse@maths.ox.ac.uk
Steve	Brewer	University of Southampton	s.brewer@soton.ac.uk
Hong	Bui	University of Southampton	Hong.Bui@soton.ac.uk
Gunnar	Carlsson	Stanford University	Gunnar@math.stanford.edu
Fai	Cheng	Lloyd's Register Group	fai.cheng@lr.org
Ying	Cheong	University of Southampton	kc5@soton.ac.uk
Matthew	Cobain	University of Southampton	mrdc1g10@soton.ac.uk
Dave	Coplin	Microsoft UK	dcoplin@microsoft.com
David	Culliford	University of Southampton	djc202@soton.ac.uk
Conrad	D'Souza	University of Southampton	conraddsouza@btinternet.com
Srinandan	Dasmahapatra	University of Southampton	sd@ecs.soton.ac.uk
Michael	Farber	Queen Mary University of London	m.farber@qmul.ac.uk
Jon	Forster	University of Southampton	J.J.Forster@soton.ac.uk
Jeremy	Frey	University of Southampton	j.g.frey@soton.ac.uk
Bogdan	Gabrys	Bournemouth University	BGabrys@bournemouth.ac.uk
Jelena	Gbric	University of Southampton	j.grbic@soton.ac.uk
Arthur	Gretton	University College London	arthur.gretton@gmail.com
Peter	Grindrod	University of Oxford	grindrod@maths.ox.ac.uk
Richard	Guerrero	University of Southampton	rel1a13@soton.ac.uk
Richard	Luduena	University of Southampton	rel1a13@soton.ac.uk
Yike	Guo	Imperial College London	y.guo@imperial.ac.uk
Stephen	Haben	University of Oxford	haben@maths.ox.ac.uk
Wendy	Hall	University of Southampton	wh@soton.ac.uk
Frank	Han	University of Southampton	frank.han@soton.ac.uk
David	Hand	Imperial College London	d.j.hand@imperial.ac.uk
Sam	Hepenstal	Hampshire County Council	Sam.Hepenstal@hants.gov.uk
Rebecca	Hoyle	University of Southampton	R.B.Hoyle@soton.ac.uk
Eugenie	Hunsicker	Loughborough University	E.Hunsicker@lboro.ac.uk
Goran	Isakovski	SITA	goran.isakovski@sita.aero
Benjamin	Jesty	University of Southampton	bj1g11@soton.ac.uk
Doha	Kattan	University of Southampton	dak1g13@soton.ac.uk
Nicola	Knight	University of Southampton	nk1g09@soton.ac.uk
Peter	Kropholler	University of Southampton	p.h.kropholler@soton.ac.uk
Vitalyi	Kurlin	Durham University & Microsoft Research Cambridge	vitaliy.kurlin@durham.ac.uk
Siu Por	Lam	Capital Normal University	siuporlam@yahoo.co.uk
Dale	Lane	IBM UK	dale.lane@gmail.com
Phuoc (Philip)	Le	University of Southampton	phl1r12@soton.ac.uk
Ray	Liu	Takeda Pharmaceutical Inc	ray.liu@takeda.com
Frank	Longford	University of Southampton	fl7g13@soton.ac.uk
Sigourney	Luz	University College London	s.luz@ucl.ac.uk
Tiejun	Ma	University of Southampton	tiejun.ma@soton.ac.uk

Ben	MacArthur	University of Southampton	bdm@soton.ac.uk
Mahesan	Niranjan	University of Southampton	mn@ecs.soton.ac.uk
Theodore	Manassis	Office for National Statistics	tmanassis@gmail.com
Enrique	Marquez	University of Southampton	esm1g14@soton.ac.uk
David	Matthews	University of Southampton	dm1x07@soton.ac.uk
Timothy	Metcalf	University College London	t.metcalf@ucl.ac.uk
Luis	Montesdeoca	University of Southampton	ljm1e14@soton.ac.uk
Enrico	Mossotto	University of Southampton	em1c14@soton.ac.uk
Katarzyna	Musial-Gabrys	Bournemouth University	kmusialgabrys@bournemouth.ac.uk
Jane	Naylor	Office for National Statistics	jane.naylor@ons.gsi.gov.uk
Tri-Dung	Nguyen	University of Southampton	T.D.Nguyen@soton.ac.uk
Sofia	Olhede	University College London	s.olhede@ucl.ac.uk
Erika	Parkinson	University of Southampton	epp@soton.ac.uk
Boonyarat	Phadermrod	University of Southampton	bp6g12@soton.ac.uk
Chris	Phethean	University of Southampton	cjp106@soton.ac.uk
Sonya	Ridden	University of Southampton	sonya.ridden@gmail.com
Alex	Rogers	Oxford University	alex.rogers@cs.ox.ac.uk
Alistair	Sackley	Hampshire County Council	alistair.sackley@hants.gov.uk
Ruben	Sanchez-Garcia	University of Southampton	R.Sanchez-Garcia@soton.ac.uk
Marc	Scott	University of Southampton	m.scott@soton.ac.uk
Paul	Skipp	University of Southampton	pjss@soton.ac.uk
Paul	Smith	University of Southampton	p.a.smith@soton.ac.uk
Rosanna	Smith	University of Southampton	r.c.smith@soton.ac.uk
Peter W	Smith	University of Southampton	pws@soton.ac.uk
Christopher	Smithers	Durham University	christopher.smithers@dur.ac.uk
Iva	Spakulova	Office for National Statistics	iva.spakulova@gmail.com
Steven	Squires	University of Southampton	ses2g14@soton.ac.uk
Benjamin	Stanley	IBM UK	benstanley@live.co.uk
Doroteya	Staykova	University of Southampton	D.Staykova@soton.ac.uk
Fabio	Strazzeri	University of Southampton	strazzeri.fabio@gmail.com
James	Strudwick	University of Southampton	jstrudwick@hotmail.com
Patrick	Stumpf	University of Southampton	PS.Stumpf@soton.ac.uk
Janette	Thompson	University of Southampton	J.L.Thompson@soton.ac.uk
Ulrike	Tillman	University of Oxford	Ulrike.Tillmann@maths.ox.ac.uk
Anne	Trefethen	University of Oxford	anne.trefethen@oerc.ox.ac.uk
Michael	West	University of Southampton	michael.west@soton.ac.uk
Paul	White	University of Southampton	prw@isvr.soton.ac.uk
Frank	Wood	University of Oxford	fwood@robots.ox.ac.uk
Tricia	Worby	University of Southampton	pw7@soton.ac.uk