

# **IDA 2017** October 26 – 28, London, UK

The Sixteenth International Symposium on Intelligent Data Analysis

# Welcome to IDA 2017

Dear Colleagues,

I am delighted to welcome you to the Sixteenth International Symposium on Intelligent Data Analysis, being hosted this year by Birkbeck, University of London. It was also at Birkbeck that the first independent IDA symposium occurred twenty years ago.

IDA 2017 features an exciting programme of three invited talks, six regular research sessions and a poster session. There is also a *First Look Session* which provides a forum for potentially ground-breaking research that is not yet mature for archival dissemination.

The symposium features invited talks from three distinguished speakers: Paul Cohen, A. Aldo Faisal and Niels Peek. I am very grateful to all our invited speakers for accepting our offer to present their work. I am also grateful to our sponsors KNIME, Springer and World Scientific for their generous support.

I hope you enjoy IDA 2017 and also have time to explore this part of London.



David Weston, General Chair

We are grateful to:



for sponsoring the Frontier Prize.



for sponsoring the Reception.



for sponsoring the Poster Prize.

### Thursday 26th October

- 08:30 Registration desk opens
- 09:15 Welcome from Master of Birkbeck
- 09:30 Invited talk: Opportunities and Challenges of Learning Health Systems by Professor Niels Peek

Chair: Allan Tucker

- 10:30 Coffee break
- 11:00 Contributed Session 1

Chair: Jaakko Hollmen

# Adapting supervised classification algorithms to arbitrary weak label scenarios.

Miquel Perelló-Nieto, Raúl Santos-Rodríguez and Jesús Cid-Sueiro

#### Searching for Spatio-Temporal-Keyword Patterns in Semantic Trajectories

Fragkiskos Gryllakis, Nikos Pelekis, Christos Doulkeridis, Stylianos Sideridis and Yannis Theodoridis

# Bucket Selection: A Model-Independent Diverse Selection Strategy for Widening

Alexander Fillbrunn, Leonard Wörteler, Michael Grossniklaus and Michael Berthold

- 12:30 Lunch
- 13:30 Contributed Session 2

Chair: Matthijs van Leeuwen

#### Improving Cold-Start Recommendations with Social-Media Trends and Reputations

João Santos, Filipa Peleja, Flavio Martins and João Magalhaes

#### The Actors of History: Narrative Network Analysis Reveals the Institutions of Power in British Society Between 1800-1950

Thomas Lansdall-Welfare, Saatviga Sudhahar, James Thompson and Nello Cristianini

# Interactive Pattern Sampling for Characterizing Unlabeled Data

Arnaud Giacometti and Arnaud Soulet

- 15:00 Coffee break
- 15:30 Contributed Session 3

Chair: Panagiotis Papapetrou

**Improving Chairlift Security with Deep Learning** *Kevin Bascol, Remi Emonet, Elisa Fromont and Raluca Debusschere* 

# Identifying novel features from specimen data for the prediction of valuable field trips

Nicky Nicolson and Allan Tucker

# Towards automatic evaluation of asphalt irregularity using smartphone's sensors

Vinicius Souza, Everton Cherman, Rafael Rossi and Rafael Souza

17:00 End of Sessions

#### 18:00 - 20:00 Reception – Sponsored by Springer

Clore Management Centre Torrington Square Bloomsbury, WC1E 7JL

### Friday 27th October

- 09:00 Registration desk opens
- 09:30 Invited talk: Reverse engineering human decision making from high resolution analysis of behaviour by Dr A. Aldo Faisal

Chair: David Weston

- 10:30 Coffee break
- 11:00 Contributed Session 4

Chair: Elizabeth Bradley Biclustering Multivariate Time Series Ricardo Cachucho, Siegfried Nijssen and Arno Knobbe

Freudian Slips: Analysing the Internal Representations of a Neural Network from its Mistakes Sen Jia, Thomas Lansdall-Welfare and Nello Cristianini

Visualization of Topic-Sentiment Dynamics in Crowdfunding Projects Rafael Carmo, Soong Moon Kang and Ricardo Silva

- 12:30 Lunch
- 13:30 Contributed Session 5

Chair: João Magalhaes **Predictive clustering trees for hierarchical multi-target regression** Vanja Mileski, Sašo Džeroski and Dragi Kocev

# ABIDE: Querying Time-Evolving Sequences of Temporal Intervals

Orestis Kostakis and Panagiotis Papapetrou

# A Dynamic Adaptive Questionnaire for Improved Disease Diagnostics

Xiaowei Kortum, Frank Klawonn, Lorenz Grigull and Werner Lechner

- 15:00 Coffee break
- 15:30 Regular poster session (including spotlight talks)
- 17:00 End of Sessions
- 19:00 Banquet Number Twelve Restaurant & Bar

12 Upper Woburn Place WC1H 0HX

### Saturday 28th October

- 09:00 Registration desk opens
- 09:30 Invited talk: Mining for Causal Results by Professor Paul Cohen

Chair: Niall Adams

- 10:30 Coffee break
- 11:00 Contributed Session 6

Chair: Stephen Swift

# Computational Topology Techniques for Characterizing Time-Series Data

Nicole Sanderson, Elliot Shugerman, Samantha Molnar, James Meiss and Elizabeth Bradley

Estimating Sequence Similarity from Contig Sets Petr Ryšavý and Filip Zelezny

#### **Hierarchical Novelty Detection**

Paolo Simeone, Tijl De Bie, Jefrey Lijffijt, Matt McVicar and Raúl Santos-Rodríguez

- 12:30 Lunch
- 13:30 First Look Session (Horizon Talks)

Chair: Elisa Fromont

Boolean Matrix Factorisation for scalable detection of overlapping network communities

Tammo Rukat and Christopher Yau

#### Dynamic Recommendations for Finding Local Expert using Lazy Random Walk with Application to Shopping Places Recommender Systems

Diyah Puspitaningrum, Julio Fernando, Edo Afriando and Ferzha Putra Utama

Towards Explanatory Data Analysis Matthijs van Leeuwen Federated Searching with Natural Language Interface to Database Richard Skeggs, Stasha Lauria and Stephen Swift

- 15:10 Prize award ceremony + closing remarks
- 15:30 16:00 Coffee
- 15:30 IDA council meeting (council only)
- 18:00 Farewell social event: TBA

# **Invited Speakers**

#### Professor Paul Cohen University of Pittsburgh

#### **Mining for Causal Results**

Science proceeds from description and classification to prediction, then strives for explanation. Classification and prediction do not require causal knowledge, explanation does. Thus, the most valuable scientific results establish cause-effect relationships. While these relationships may be implicit in data, they are explicitly stated in scientific publications. Extracting causal relationships from data is difficult, whereas extracting explicit causal assertions from scientific papers is relatively easy; which makes today's preoccupation with Big Data at least puzzling and perhaps wrong-headed. Hints from DARPA's Big Mechanism program suggest that domain-specific causal knowledge in scientific papers can be mined by machines using domain-independent methods, which would make these the methods of choice for scientists who prefer causal assertions to vast amounts of data that might or might not harbor causal assertions. A "third way" envisions data as a source of causal assertions that scientists might otherwise miss, and as evidence pertaining causal assertions mined from scientific literature. Of course, shifting the unit of analysis from data to causal assertions raises many research issues, including the problem of integrating published assertions, which tend to be fragmentary, into larger models or theories; and the need to maintain the experimental contexts of these assertions; and the likelihood that at least some assertions are wrong. The talk will illustrate these points with concrete examples of mining causal assertions about cancer biology.

Paul Cohen is founding dean of the School of Computing and Information at the University of Pittsburgh. Prior to that he managed programs at the Defense Advanced Research Projects Agency (DARPA), part of the US Department of Defense, while on leave from the University of Arizona. At Arizona, Cohen served as head of Computer Science and founding director of the university's School of Information: Science, Technology and Arts. He holds a PhD in Computer Science and Psychology from Stanford University, a Master of Science degree in Psychology from the University of California, Los Angeles, and a Bachelor of Science degree in Psychology from the University of California, San Diego.

Cohen's research interests span artificial intelligence and include machine learning, language, vision, semantic technology, data analysis, information theory and education informatics. His programs at DARPA are concerned with technologies to help humans model and manage very complicated systems, such as cancer molecular signalling pathways and the food-water-energy nexus.

### **Invited Speakers**

#### Dr A. Aldo Faisal Imperial College London Reverse engineering human decision making from high resolution analysis of behaviour

Our research focuses the computational analysis of human behaviour using data-driven analysis and prediction, including the development of methods & algorithms to move from Data to Knowledge. Our key Goals are to predict human behaviour from ubiquitous sensors & digital data and so to predict and evaluate human performance, as well as inferring internal or cognitive state of individuals from behavioural dynamics. We will illustrate these goals with current examples in our Human Ethome Project as well as supporting medical decision making in the clinic with pervasive sensing.

Dr Aldo Faisal is Associate Professor for Neurotechnology at Imperial College London (Dept. of Bioengineering & Dept. of Computing) and Director of the Behaviour Analytics Lab at the Data Science Institute (London). He is Associate Group Head at the MRC London Institute for Medical Sciences and affiliated faculty at the Gatsby Computational Neuroscience Unit (University College London). In 2009 he was appointed to Imperial College London to develop a human-based Neurotechnology research area, where he established his lab. Dr Faisal's combines cross-disciplinary computational and experimental approaches to investigate how the brain and its neural circuits are designed to learn and control goal-directed movements. The neuroscientific findings enable the targeted development of novel technology for clinical and research applications (Neurotechnology) for a variety of neurological/motor disorders and amputees. Aldo serves on the editorial board of Nature Scientific Data and PLoS Computational Biology, and at international conferences (general/program/area chair) such as Neurotechnix, various IEEE conferences and NIPS. He was recently elected to the World Economic Forum's Council on the future of Neurotechnologies and Brain Science.

# **Invited Speakers**

#### Professor Niels Peek University of Manchester Opportunities and Challenges of Learning Health Systems

Health systems worldwide are under pressure to deliver better care for more people from fewer resources. The global economic crisis has shrunk the resources available for healthcare but the growth in demand for care services continues unabated. "Learning Health Systems" is a novel health informatics paradigm that blends quality improvement methods with data science. The goal is to create integrated healthcare systems which harness the power of data and analytics to learn from every patient, and feed the knowledge of "what works best" back to clinicians, patients, and other stakeholders of the healthcare system, to create cycles of rapid improvement. In this talk we dissect the new paradigm and explore its opportunities and challenges for data scientists.

Niels Peek (MSc, PhD) is a Professor in Health Informatics at the University of Manchester. With a background in Computer Science and Artificial Intelligence, his research focuses on data-driven informatics methods for healthcare quality improvement, data mining for healthcare, predictive models, and clinical computerised decision support. He is director of the Greater Manchester Connected Health City, which is part of the B#20M "Health North" investment to establish a learning health system in the North of England. Professor. Peek has co-authored more than 150 peer-reviewed, scientific publications. Previously based at the University of Amsterdam, the Netherlands, he led the "CARDSS" initiative, a collaboration between academic partners. professional and patient organisations in cardiac rehabilitation, and industry partners which led to the introduction of computerised decision support in 40 Dutch hospitals and national quality standards for cardiac rehabilitation. He is the President of the European Society of Artificial Intelligence in Medicine, editorial board member of the Journal of the American Medical Informatics Association and Artificial Intelligence in Medicine, and co-chair of the Scientific Programme Committee of MEDINFO-2017, the 16th World Congress on Health and Biomedical Informatics.

# **Social Events**

#### Thursday 26<sup>th</sup> October

#### Reception 18:00 - 20:00

A drinks reception sponsored by Springer will be held on the first evening of the symposium. The venue is the Clore Management Centre, Torrington Square, Bloomsbury, WC1E 7JL

#### Friday 27<sup>th</sup> October

#### Banquet – Number Twelve Restaurant & Bar 19.00 – 23.00

#### 12 Upper Woburn Place WC1H 0HX



The banquet will be held at the Number Twelve Restaurant & Bar which is located across the road from Woburn House. This restaurant specialises in British Italian fusion cuisine.

Arrival Drinks: 19.00 – 19.25

**Dinner:** 19.30 – 23.00

# **Places to Visit**

Woburn House is located in the heart of Bloomsbury, an area that was made famous by the *Bloomsbury Set* who were a group of artists, writers and philosophers that lived in the area in the early 20<sup>th</sup> Century.

There are plenty of places to visit that are in easy walking distance. (Times measured from Woburn House)

**Tavistock Square Gardens –** also known as the Peace Garden contains several memorials, 1 minute.

**Wellcome Collection** – The Wellcome Trust's free museum, 5 minutes.

British Library – 8 minutes.

British Museum – 14 minutes.

If you plan to go out for a meal, **Goodge Street** (13 minutes) has plenty of restaurants many of which are reasonably priced.

#### **WiFi Network**

Delegates have WiFi access during the conference, **eduroam** is available. Alternatively, you can:

- 1. Connect to the wireless network: Woburn House Conference Centre
- 2. Ask for credentials from reception (not available in online version of this document):

#### **Online Access to conference proceedings**

Springer have arranged temporary free online access to the proceedings for conference delegates.

Follow the link on the conference website home page: http://www.dcs.bbk.ac.uk/ida2017/

# Мар



- Banquet Number Twelve Restaurant & Bar 12 Upper Woburn Place WC1H 0HX
- Woburn House Conference Centre 20-24 Tavistock Square, WC1H 9HQ
- Reception Clore Management Centre Torrington Square Bloomsbury, WC1E 7JL

# **IDA 2017**







Department of **Computer Science and** Information Systems



