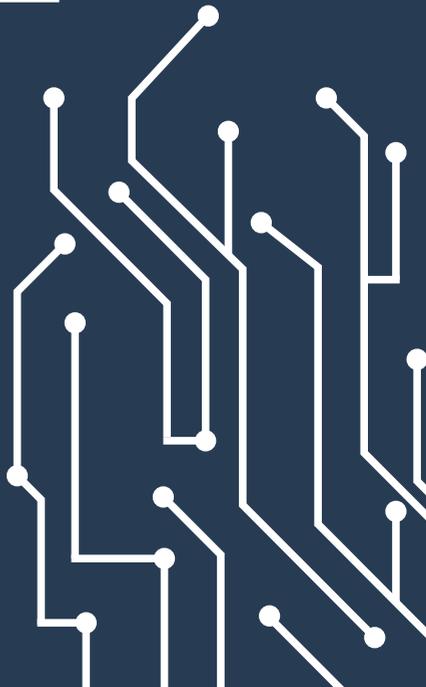


WPCCS
2017



It is my pleasure to welcome you to the Warwick Postgraduate Colloquium in Computer Science 2017 (WPCCS), to be held from this year onward in a new format. The organizers have worked very hard to reshape WPCCS into an interesting and captivating event. It is now your turn to make it better by sharing your research experiences, ideas, and visions with your peers.

What brings you all together is your passion for innovation and an inherent desire to achieve something outstanding in the field of Computer Science. Unlike most of your former colleagues you have chosen the unknown path, yet you are confident that your perseverance to face the murky facets of research and your ingenuity to solve open problems will guide you through. And they will.

You have been fortunate to explore your talent in Warwick CS, which is one of the most fertile breeding grounds for creativity like yours. Our Department is equipped with a fine team of researchers who once undertook unique challenges like yours. Not only did they succeed, but they are now here because they want to further succeed through you. Learn from them, leverage their experience and intuition, but do remain assertive by trusting your instincts. Working closely together will largely narrow the gap to your joint success.

WPCCS is an excellent opportunity to widely expose where you currently stand. Speak up and do enjoy letting yourself be challenged!

- Florin Ciucu



“In October we will be celebrating 50 years of Computer Science at Warwick”

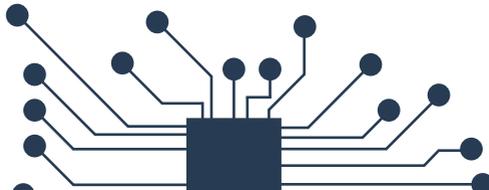
Welcome to WPCCS 2017

The postgraduate population in Computer Science continues to grow and there are now more than 100 registered PhD students in the Department. At the same time, the number of academics in Computer Science is also growing and we are pleased to be adding expertise in areas including cyber-physical systems, databases, computational modelling and dynamic graph algorithms, amongst others. Research students and staff are enjoying the benefits of the Department's association with the Center for Urban Science in New York and the Alan Turing Institute in London, as well as with the many companies and organisations that work with us.

In October we will be celebrating 50 years of Computer Science at Warwick. It is fitting therefore that our 40,000 ft² laboratory extension will open in this 50th year, marking a forward-looking transition to the new era. I often have the pleasure of meeting alumni, who are now business leaders, wealthy entrepreneurs or leading academics. They remind me that Warwick continues to be an excellent institution in which to conduct research, with some of the strongest academic departments in the country, and that Computer Science has an enviable reputation for impactful research, which remains at the forefront of change in our society.

Good luck with WPCCS 2017 and I hope you have an interesting and instructive day.

- Stephen Jarvis



TRACK 1 · OC1.07**Computer Vision**

Chair: Liam Steadman

Muhammad Shaban

Haoyi Wang

Roberto Leyva Fernandez

TRACK 2 · OC1.08**Education**

Chair: Richard Kirk

Noor Hasimah Ibrahim Teo

Ebtehal Quqandi

TRACK 3 · OC1.09**Urban Science**

Chair: David Purser

Philipp Ulbrich

Melissa Kenny

Isy Slattery

BREAK

Jamie Bayne

Ian Tu

Simon Graham

Shan Lin

Dimah Al-Fraihat

Nicole Peinelt

Nouf Almujaally

Huda Alrashidi

Vangelis Pitidis

Katherine Harris

Aseel Alturki

Corinne Muir

LUNCH**TRACK 4 · OC1.07****Foundations**

Chair: James Van Hinsbergh

Charlie Dickens

David Purser

Tejas Kulkarni

Eleanor Davies

Shenyuan Ren

TRACK 5 · OC1.08**Networking**

Chair: Richard Kirk

Betty Agbons

Farrukh Qazi

Matthew Bradbury

Jack Kirton

Peter Davies

TRACK 6 · OC1.09**Urban Analytics**

Chair: Liam Steadman

Neha Gupta

Timothy Sit

Elisabeth Titis

Alex Caton

BREAK**TRACK 7 · OC1.07****High Performance
Computing**

Chair: Matthew Bradbury

Jack Jackson

Andrew Owenson

Richard Kirk

Dom Brown

Zhenyu Li

TRACK 8 · OC1.08**Artificial Intel.
& Transport**

Chair: David Purser

Helen McKay

James Van Hinsbergh

James Marchant

James Archbold

Liam Steadman

TRACK 9 · OC1.09**Machine
Learning**

Chair: Greg Watson

Junyu Li

Zhuoer Gu

Christopher Hickey

Edward Chuah

Latifah Almuqren

POSTERS

Zakiyya Adam

Mohammed Alghamdi

Al Alharbi

Sarunkorn Chotvijit

Henry Crosby

Denys Flores Armas

Adam Gelencser

Chen Gu

Vikki Houlden

Nicholas Johnson

Martina Kluvancova

Elena Kochkina

Caroline Player

Lee Prangnell

Ayman Qahmash

John Rahilly

Nataliya Tkachenko

David Truby

Adam Tsakalidis

Bo Wang

Greg Watson



WHEN & WHERE

9:20 am Welcome from Prof. Stephen Jarvis in OC0.02

9:30 am Guest talk from Dr. Sara Kalvala in OC0.02 (see page 6)

10:00 am Tracks 1, 2 & 3

10:50 am Break with refreshments in foyer

11:10 am Tracks Continue

12:10 pm Lunch and posters in ground floor foyer

1:00 pm Guest talk from Vic Smith in OC0.02 (see page 7)

1:30 pm Tracks 4, 5 & 6

2:50 pm Break with refreshments in foyer

3:10 pm Guest talk from Alex Haak in OC0.02 (see page 7)

3:40 pm Tracks 7, 8 & 9

5:10 pm Prizes awarded in OC0.02 (see page 8)



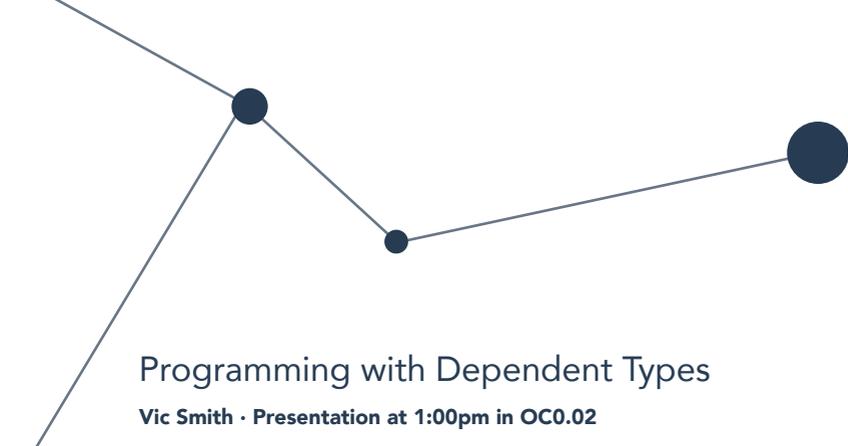
Guest Talks

Microbial communities: Understanding and exploiting the origins of cooperative behaviour

Dr. Sara Kalvala · Presentation at 9:30am in OC0.02

Micro-organisms are all around us - usually forming rich multi-species communities where hundreds of thousands of individuals interact, whether to cooperate, compete, or feed on each other. These communities of micro-organisms affect many aspects of our day-to-day life, and are the focus of experimental research by biologists, biochemists, bio-engineers and environmentalists. There is also significant interest by mathematicians, computer scientists and even social scientists and philosophers in studying microbial communities, as concepts of cooperation and competition are relevant not only at the micro-organism level but everywhere - between nations, between animals, between mobile phone users, and so on. A Multi-Agent System is an abstract, computational model of how individuals interact to form communities, and the use of this abstraction allows research as well as application in the development of nature-inspired computational tools.

In this session Sara will introduce the general concept of a Multi-Agent System, from a computational perspective. She will then discuss how she uses abstract models to understand more about how microbes organize themselves into colonies and work together, and on the flip side how this understanding of how microbes work together inspires new technologies.



Programming with Dependent Types

Vic Smith · Presentation at 1:00pm in OC0.02

This talk explores the world of dependent types, primarily in the Idris language. Dependent types allow greater safety of code in a practical way, with features like proofs instead of tests, and guarantees that your programs will terminate. But they also offer much more: like generic programming, and type driven development with compiler assisted coding — which sometimes involves the compiler writing the entire program for you.

Innovation or Industry

Alex Haak · Presentation at 3:10pm in OC0.02

A look at the mutual exclusivity of innovation and industry. Does working for a company preclude you from working on the cutting edge?

At Improbable, we want to be living proof that this is not the case. We have the, somewhat ambitious, mission of using new research to transform understanding of complex problems and systems. Our tech lets organisations use large scale agent-based simulation to gain insight and help decision making. In common English, this translates to: “We do some pretty cool stuff”. We explore, research and experiment, all whilst delivering real value and making money.

Featuring rocket science, the matrix and other unrelated concepts.





Prizes

WPCCS aims to engage attendees with their colleagues' research and with research in the wider community. For this purpose, the Department of Computer Science and the Research Student Skills Programme have kindly agreed to sponsor the event.

For today's posters and presentations, 10 prizes will be awarded by the Programme Committee. An Amazon voucher will be awarded per track for the presentation that best informs and educates with engaging deliverance and another to the most insightful poster. One of these prizes will be awarded as 'Best in Colloquium'.

TRACK 1 · OC1.07 · 10:00am

Computer Vision



Introduction by **Dr. Abhir Bhalerao**

Chaired by Liam Steadman

Context-aware Patch Classification in Whole Slide Tissue Images

Muhammad Shaban · Presentation at 10:05am

Convolutional Neural Networks (CNN) models are best suited to many image classification tasks. However, this is not true for classification of gigapixel resolution Whole Slide Tissue Images (WSI). Processing a WSI as whole through a CNN is computationally unfeasible. The only plausible option is patch-level classification of WSI into different classes (tumour and benign). However, classification of some tumour classes requires large contextual information which cannot be captured by a single patch. We proposed a context aware stacked CNN based patch classifier that will consider the feature of neighbouring patches while classifying a given First network takes a larger patch, which should consist on multiple small patch, and convert its sub-patches into high dimensional feature space and then second network take those features and predict the class of each sub-patch using context information. This approach will be useful to improve the accuracy of simple patch based classifier and WSI classification.

Face Recognition Against Ageing with Deep Learning

Haoyi Wang · Presentation at 10:20am

Missing Children is a critical issue that has troubled thousands and even millions of families around the world. By implementing a face recognition system that is able to against ageing, we can recognise and locate those missing children even years or decades later after the tragedies happened. Besides, deep learning is one of today's trending topics. In the presentation, I will firstly review what is deep learning and some state-of-the-art convolutional neural network structures, followed by some recent papers on age estimation and the ageing problem that involved deep learning. Then, my proposed method and current progress are discussed.

Detecting Realistic Events in Surveillance Videos

Roberto Leyva Fernandez · Presentation at 10:35am

There are millions of cameras fitted in public scenarios all over the world, more than ever it is important to help the monitors to detect certain events as soon as they happen. Such events (e.g. robberies, traffic accidents, vandalism, etc.) must be detected in real time to properly take an action immediately. Still many important aspects must be addressed to achieve competitive performance for practical applications purposes (e.g. reduce false alarms triggers, computational complexity, hardware demands, high descriptive feature extraction, etc.). In this talk we present more in detail: (I) an end-to-end framework to detect such events in real time with competitive results. (II) The surveillance material employed to detect dangerous events and privacy concerns (III) open challenges and important aspects that would make automatic video surveillance more feasible under our scope.

Towards Stereo Vision ConvNets in Real-Time

Jamie Bayne · Presentation at 11:10am

The topic of stereo depth perception is one of the oldest in computer vision, but still an active area of research. The problem is computationally expensive, and current work centres around making algorithms accurate and fast enough for autonomous vehicles. More recently, deep learning solutions have shown state-of-the-art accuracy, but are too slow to meet real-time requirements. This presentation will briefly explain deep learning's place amongst more traditional approaches, and then examine how these networks are structured and what can be done to make them faster.

Deep Passenger State Monitoring

Ian Tu · Presentation at 11:25am

The advent of autonomous and semi-autonomous vehicles has meant passengers now play a more significant role in the safety and comfort of vehicle journeys. This talk shall present a deep learning method to monitor and classify passenger state with in-vehicle camera data. The passenger state detection method uses a convolutional neural network in combination with viewpoint warping using planar homography. This method allows data which usually cannot be included at the training stage to be effectively used for re-training, and data re-purposed from driver monitoring to occupant state classification. The viewpoint normalisation and augmentation also allows the trained model to be re-trained with additional data to work between vehicle types. To evaluate the robustness of the proposed method, data was collected in two different vehicles at three different viewpoints, and used to demonstrate that viewpoint is a significant factor influencing accuracy.

Hierarchical convolutional neural network for segmentation of tumour rich areas in breast cancer histology images

Simon Graham · Presentation at 11:40am

The rise in digital pathology has been matched with an increase in interest for automated tumour segmentation in hematoxylin & eosin (H&E) stained histology images. Automated detection of tumour rich areas holds significant clinical relevance to pathologists, due to a reduction in diagnosis subjectivity and time implementing the segmentation. We propose a hierarchical convolutional neural network (HCNN) for detection of tumour regions in whole slide images (WSIs) of breast histology slides. The proposed network architecture consists of down-sampling layers that capture the context of the image, followed by up-sampling layers that localise the region of the tumour. Experimental results show that the proposed network performs favourably over conventional convolutional neural networks (CNNs).

End-to-End Latent Relationship Learning of Mid-Level Deep Features for Person Re-ID

Shan Lin · Presentation at 11:55am

In person re-identification, human attributes such as long hair, blue shirt, etc. are distinctive mid-level features which are robust to variation of illumination, misalignment, and viewpoints. However, manual labeling attributes is a costly task. In our proposed method, we utilize the convolutional layer output from the ImageNet pre-trained network as the replacement for manually labeled human attributes. We propose an end-to-end deep mid-level feature learning approach to (1) automatically select the best features for feature representation in a data-driven approach (2) The inter-personal and intra-personal correspondence relationships of these mid-level features will also be adaptively learned in our proposed network. Our method achieved 79% rank-1 accuracy to CUHK03 and 74% rank-1 accuracy in CUHK01 which is improved over 20% comparing to Improve_DL and FPNN. Our model can reach over 90% accuracy after rank-5. Moreover, our proposed structure show promising performance in the cross-dataset scenario. When our network train from CUHK03, it can achieve 64% rank-1 accuracy in CUHK01.

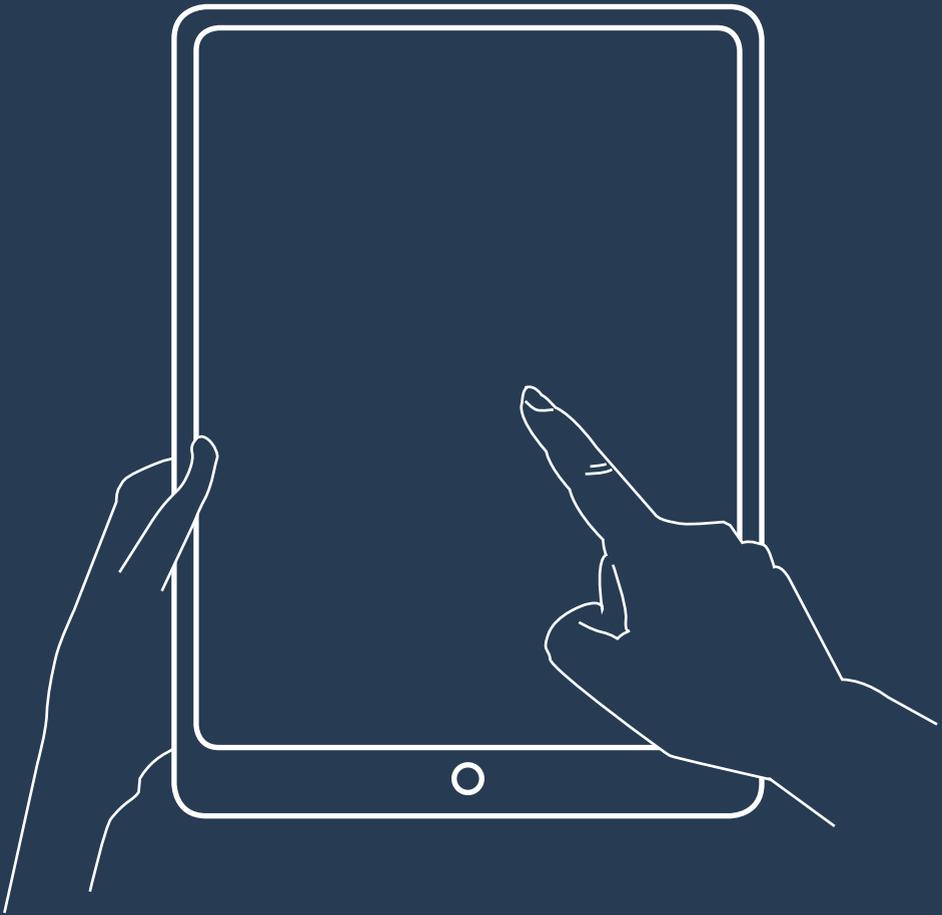
Person Re-Identification using Partial Least Squares Appearance Modelling

Greg Watson · Poster presented in foyer

Person Re-Identification is an important task in surveillance and security systems. Whilst most methods work by extracting features from the entire image, the best methods improve performance by prioritising features from foreground regions during the feature extraction stage. In this paper, we propose the use of a Partial Least Squares Regression model to predict the skeleton of a person, allowing us to prioritise features from a person's limbs rather than from the background. Once the foreground area has been identified, we use the LOMO and Salient Colour Names features. We then use the XQDA Distance Metric Learning method to compute the distance between each of the feature vectors. Experiments on VIPeR, QMUL GRID and CUHK03 data sets demonstrate significant improvements against state-of-the-art.

TRACK 2 · OC1.08 · 10:00am

Education



Introduction by Dr. Steve Matthews

Chaired by Richard Kirk

Generating Question Templates from Course Ontology Elements for Automatic Question Generation

Noor Hasimah Ibrahim Teo · Presentation at 10:05am

Ontology-based question generation has been widely studied recently. This kind of research aims to benefit instructors by providing support and intelligent assistance for the automatic generation of questions. However, existing ontologies are not designed mainly for this purpose, and the concern is that ontology will not be competent enough to act as a semantic source for the question generation process. Therefore, the aim of this work is to validate how well the elements represented in course ontology can be used for the purpose of automatic question generation. In this work, we choose to validate Operating System ontologies and identify related question sources from textbooks on this subject as competency questions. The result shows that the evaluated ontologies can provide knowledge for generating useful assessment questions. Furthermore, the list of categorized question templates and their variations are generated using a strategy based on the validated ontology.

Mobile Augmented Reality in Educational Environments

Ebtehal Quqandi · Presentation at 10:20am

We live today in a smart world. There are a plethora of smart devices surrounding us, which we use in our daily lives. Additionally, we are easily able to access the Internet via those devices whenever for whatever and wherever we need. This rapid development of technology creates rich resources that could lead to change the infrastructure of traditional learning methods to intelligent learning by deploying new technology in a smart learning environment. One of the new technology is Augmented Reality (AR), which creates virtual extra layers on the physical object. It allows bringing digital information into the real environment by blending those two worlds together. AR is offering new learning opportunities, integrated with Mobile applications. The Smartphone devices can be utilized as AR tools in the way of interactive learning. Despite, there is an acknowledgment that AR application has a positive impact on learning outcomes, there is still a lack of guideline for integrating AR technology with learning theories and lack of guideline evaluating for AR educational applications. My research will address the gap of integrating AR with mobile learning based on learning theories.

An Evaluation of E-learning from Users' Perspective: The Case of Moodle at the University of Warwick

Dimah Al-Fraihat · Presentation at 11:10am

The aim of this research is to investigate the factors that contribute to the success of e-learning environment from users' perspective. The aim of the research drives the development of a conceptual framework for e-learning success factors. It is hoped that identifying the factors that influence e-learning in terms of pedagogy dimension will be a contribution to the field of e-learning generally and in particular the issues related to the Moodle. In addition, the framework can be used as a reference to give insights for higher education institutes through understanding the factors that influence an e-learning environment.

Community-based Question Answering for Massively Open Online Courses

Nicole Peinelt · Presentation at 11:25am

In a time of ongoing knowledge expansion, the need to continuously train and retrain the working population is becoming increasingly evident. With substantial cost pressure on traditional education systems, open educational resources such as Massively Open Online Courses (MOOCs) provide a promising alternative to meet this demand on a large scale. Despite several benefits over traditional education systems such as flexibility, scalability and democratisation of educational resources, studies have highlighted current limitations of MOOCs, such as low completion rates and a lack of feedback. This project aims at applying current question answering methodologies to automatically provide feedback to students' questions in MOOC forums and reduce the workload for human facilitators.

A new model for the adoption of web-based knowledge sharing system amongst academics in Saudi Arabian higher education institutions

Nouf Almujaally · Presentation at 11:40am

Knowledge sharing has become a significant source of success in Knowledge Management (KM). In many organizations, knowledge management is often inadequate when it comes to web-based knowledge sharing, particularly among academics who work in Saudi universities. To ensure that a knowledge sharing system can be well implemented and used when communicating internally in an academic context, there is a need to know why academics accept or reject the use of web-based knowledge sharing systems. Therefore, the aim of this research is to determine the factors that affect academics' behaviours toward using web technology to share knowledge in Saudi Arabian higher education institutions. Furthermore, a web-based knowledge sharing adoption model is constructed based on the factors which already exist in technology acceptance theories, such as the Unified Theory of Acceptance and Use of Technology (UTAUT) and Task-Technology Fit (TTF), as well as other factors which are explored in knowledge sharing literature reviews to enrich the proposed model. Then, the model will be edited and refined using mixed method approach. The future work will expand the model and evaluate it to ensure that it fits the academics' needs.

Reflective Writing in Computer Science

Huda Alrashidi · Presentation at 11:55am

Providing opportunities for students to express their reflective thinking (RT) in writing is an important educational goal. Considering the amount of text provided by online learning system and applying reflection on such text would offer a great benefit of research. In this work, a hybrid approach based on keywords selection, rules inferences, and machine learning is presented to detect reliable reflective writing (RW) in the text. The text is collected from computer science writings to provide novelty work and research. This is based on analysing each text segment from learner writings.

The most prominent of RW models are investigated to gather the main categories of RW analysis. Several approaches are employed to apply a reliable detection of RW for each category. It was founded that humans outperform the automated methods regarding the level of reliability. The introduced approach can be used to automatically detect reflection in text providing a technology tool for teachers or learners to consider and analyse their personal practices and others' perspectives.

Investigating high-achieving students' code-writing abilities through the SOLO taxonomy

Ayman Qahmash · Poster presented in foyer

Computer Science Educationalists have implemented educational taxonomies which enhance the pedagogy for introductory programming modules. The SOLO taxonomy has been applied to measure students' cognitive abilities in programming by classifying students' exam answers. However, SOLO provides a generic framework that can be applied in different disciplines, including Computer Science, and this can lead to ambiguity and inconsistent classification. In this paper, we investigate high-achieving students' coding abilities and whether they tend to manifest specific SOLO categories. We address the challenges of interpreting SOLO and the limitations of code-writing problems by analysing three specific programming problems (Array Creation, Linear Search and Recursion) and solutions to those problems presented by a group of nine students. Results for the first programming problem show that six students' responses fell into the highest possible category (Multistructural) and the remaining three were categorised in the second highest category (Unistructural). For the second problem, eight students' responses fell into the Multistructural category, while only one response was categorised as Unistructural. For the third problem, two students provided Multistructural solutions and five students' solutions were Unistructural, but two further students showed a lack of understanding program constructs in their solutions, which were then categorised as Prestructural.

TRACK 3 · OC1.09 · 10:00am

Urban Science



Introduction by Prof. Rob Procter

Chaired by David Purser

How does resilience-thinking make urban critical infrastructure management more effective?

Philipp Ulbrich · Presentation at 10:05am

Increased urban critical infrastructure (CI) interdependencies and a rising unpredictability of the operating environment have moved the urban CI governance debate from traditional risk management towards the enhancement of adaptive capacity, understood as the ability to reconfigure in anticipation and reaction to known and unknown threats. This research proposes an analysis of how new modes of governance emerge from different interpretations of “critical infrastructure” and “resilience” and how these are reflected in changing organisational forms. Adopting mixed methods approach, the research will apply critical discourse analysis to identify changes in the organisation of urban CI governance, specifically by observing the translation of discourses into urban CI policies and interventions that engage a wider range of stakeholder and citizens. These insights will be used to create key components for an agent-based urban CI governance model to complement the industry partner’s decision-making platform.

Urbanisation in the Arctic: Planning for Development and Change

Melissa Kenny · Presentation at 10:20am

Development and urbanisation of the Arctic is likely to become a central feature of the region. As the natural resource extraction industry is set to grow, alongside the impacts of climate change, the landscape and the livelihoods of those who live there is likely to be subject to extreme changes in the near future. Urban areas will become the focal points for development in the region, developing alongside potential industrial growth. Increased levels of accessibility due to the impacts of climate change, as well as new opportunities for resource extraction, trade routes and rural-urban migration mean that urban development the Arctic is becoming increasingly significant, economically, politically and socially. These opportunities also come with threats and consequences; remote regions are particularly environmentally vulnerable to human influence. Furthermore, cost, human adaptability and possible resilience interventions all present challenges. These complex challenges need to be anticipated and consequently well planned for and managed to ensure that the almost inevitable urbanisation of the Arctic is sustainable, manageable and inclusive at all scales.

Global city quality of life and the built environment

Isy Slattery · Presentation at 10:35am

The quality of living in world cities has never been a more important topic than today when it is accepted that almost 70 percent of the world's population will be living in cities by 2030. Whilst there is a well-established global industry involved in measuring and assessing large cities for indicators of quality of life or associated indexes of happiness or well-being, this work tends to be done at an aggregated city scale and thus lacks a vital local component. In this research I will mobilise the idea of place-making in order to fill this gap in quality of life assessment and in particular how the built environment can mediate the relationship between how citizens feel about the city. Place making in this context provides the catalyst to enable decision makers and citizens to have a say in shaping and changing their built environment to enhance the liveability of their cities.

The importance of geohazards for urban resilience: A study of Thessaloniki, Greece and its participation in the 100 Resilient Cities Network

Vangelis Pitidis · Presentation at 11:10am

This study aims to illuminate the importance of geology's role into urban resilience policies by mapping out Thessaloniki's most vulnerable areas to earthquakes and surface flooding and in performing an integrated spatial analysis. Thessaloniki, Greece adopted a Resilience Agenda in 2015 by participating in the Rockefeller 100 Resilient Cities network. The geological heterogeneity of the city's subsurface and the anthropogenic impact on the environment as a result of 2,300 years of human settlement, have contributed to its high susceptibility to geohazards, such as the earthquakes and surface flooding. Focusing on specific neighbourhoods of Thessaloniki's metropolitan area and building on the city's recently published Resilience Strategy (2017) we unveil how the different components of the city's resilience potential intermingle and manifest. The study concludes by analysing targeted urban design interventions that are prove to be extremely helpful in assisting local governments to confront the impact geohazards, as well as decreasing the level of vulnerability and the extent of exposure of affected groups.

Use of Geo-Resources for Sustainable and Resilient Urban Design in Cities

Katherine Harris · Presentation at 11:25am

The general trend in cities around the world is that urbanization is on the rise. As a result of this, cities are becoming increasingly reliant on geo-resources to support their everyday services and expansion. Geo-resources are naturally occurring assets of the earth that can be harnessed to create something functional for our consumption, including; geo-materials, sub-surface space, groundwater and geothermal energy. Above ground space, ground properties (permeability, stability, etc) and topography can also be considered valuable geo-resources that have sometimes been overlooked in city development. This study is investigating how the use of geo-resources can enhance the resilience and sustainability of urban design in cities. Current perceptions and practices of utilizing geo-resources are being reviewed, as well as an examination of the current resilience and sustainability assessment methods proposed in different cities today. Ultimately, a geo-resource systems analysis tool is envisioned to propose a way forward to resilient urban design.

Increasing Complex Networks' Resilience by taking Inspiration from Food Webs

Aseel Alturki · Presentation at 11:40am

Human built complex networks around us are facilitating our lives in many ways and they are under many natural and man-made stressors. These stressors can cause cascade failures and the creation of new hidden topological vulnerabilities. Recently, research has shifted from maintaining network's sustainability to accepting that failure is natural and focus, instead, on networks resilience. Which is the system's ability to bounce back as quick as possible and minimize cascade failures.

We can learn from natural complex systems which have evolved under constant predation and environmental stress. This project focuses on getting inspiration from food webs in terms of long-term resilient growth of complex systems. And short-term dynamically rewiring to minimize cascade failures to achieve dynamic resilience in the face of unknown risks.

This naturally occurring distributed intelligence in adapting under future uncertainty may help us develop connected thinking and build both static and dynamic approaches to resilience.

Are changing management practices in urban parks and greenspaces increasing biodiversity, and are there trade-offs with amenity?

Corinne Muir · Presentation at 11:55am

Parks are an easy target for budget cuts as there are no statutory obligations for their provision or maintenance. However, they have many important roles for urban populations including recreation, healthy exercise, air pollution reduction and flood mitigation. Various techniques are being tried in the UK's cities to reduce management costs in parks, for example reduced mowing and new types of floral display. Could some cost saving exercises improve parks for people and wildlife, or does amenity suffer?

Methods: Investigating changing management practices in parks and their effects on ecosystem services. Local fieldwork results from Coventry will be compared to national data; Comparing parks mapping (from Ordnance Survey, released soon) to scenic-ness ratings (from scenicornot.datasciencelab.co.uk). This has been trialled with existing data from Scotland and London; Comparing changing management/maintenance to habitat data in Coventry; Public opinion and park usage in different types of habitat/management type from on-site surveys in Coventry and a national online survey using photographs.

A Green and Pleasant Land

John Rahilly · Poster presented in foyer

Within a context of the seemingly relentless growth of urban populations and concomitant pressures upon land use significant pressure has been placed upon undeveloped areas, such as 'green infrastructure' land. In parallel, such space has become formally recognised as having significant environmental, economic and societal value.

Although the extent of 'green infrastructure' is influenced by environmental features, the potential impact of policy decisions upon the prevalence of such is becoming recognised as key. Accordingly, this research seeks to catalogue the prevalence of 'green infrastructure' and utilise multiple data sources in order to analyse the impacts of different approaches to planning policy over time.

As an introductory study, research was conducted with which to identify the potential utility of OpenStreetMap as a resource for contemporary data. It was analysed in regards to completeness, attributional accuracy and positional accuracy. This initial analysis suggested that such data was inaccurate for the needs of the research, but could offer significant contextual value in relation to 'unofficial' spaces.

Understanding Neighborhood Dynamics Through In-situ Sensing of the Urban Environment

Nicholas Johnson · Poster presented in foyer

The rapid pace of global urbanization has intensified the need to understand the complex and dynamic interactions within urban systems in order to provide sustainable and efficient cities and to improve quality-of-life for its citizens. Study at the neighborhood scale offers a promising lens to more fully understand social, environmental and man-made systems within and across heterogeneous communities and facilitates community participation, which is necessary to fully understand problems faced by the community and connects tacit neighborhood knowledge to research outcomes. This research describes our recent pilot project to deploy novel urban sensors in Red Hook to collect and analyze quality-of-life measurements at high spatial and temporal resolution.

Bringing Cities to Life: a multilevel spatial modelling approach to analysing greenspace types, accessibility and mental wellbeing

Vikki Houlden · Poster presented in foyer

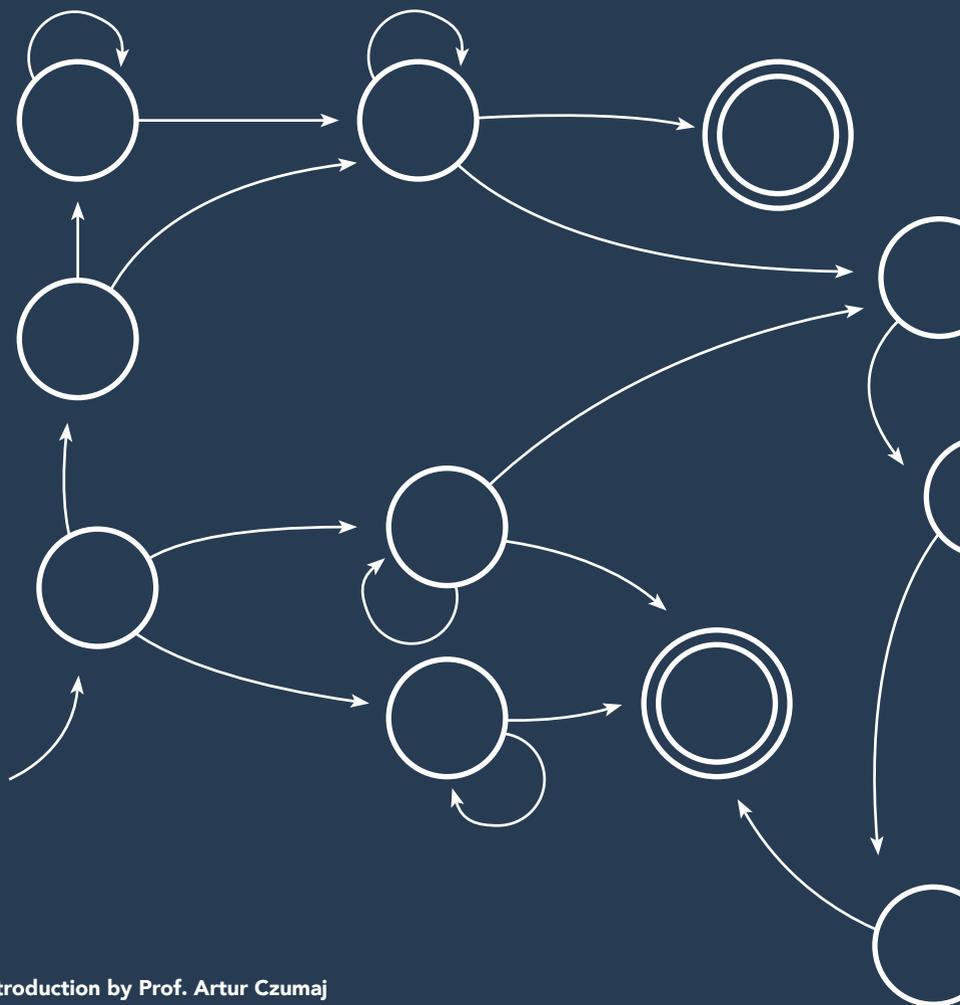
Today, over half the world's population resides in cities, creating a challenge for urban planners to accommodate new residents and a changing way of life. However, it has been suggested that mental health may be poorer in urban than rural environments, meaning that these planners have a social responsibility to try and design more healthy, positive environments. While it has long been established that people benefit from exposure to nature, it remains unclear which green features are most important for mental wellbeing and how these can be effectively designed into urban environments to create a positive mental health landscape.

Commuting and Well-Being

Zakiyya Adam · Poster presented in foyer

The commute to and from work is just part of the daily routine for a large proportion of the adult population. For most, it constitutes a short part of their day. Despite this, this journey has a significant impact on their well-being; the commute to work was found to be the worst part of a person's day, less enjoyable than housework or working (Kahneman et al, 2004). Understanding the negative impact of commuting is a relatively new field of research, and the focus of my work.

Foundations



Introduction by Prof. Artur Czumaj

Chaired by James Van Hinsbergh

Deterministic Distributed and Streaming Algorithms for Linear Algebra Problems

Charlie Dickens · Presentation at 1:35pm

Much of the work on randomized linear algebra has led to extremely efficient distributed and streaming algorithms for a number of fundamental problems such as approximate matrix multiplication, low rank approximation, and regression. In this work we ask what, if anything, is possible for deterministic algorithms. While there are notable deterministic algorithms for basic versions of these problems, such as Frobenius norm low rank approximation and least squares regression, they are scarce and do not handle more robust versions of these problems. Our goal is to initiate the study of deterministic algorithms for a wide array of linear algebra problems in the distributed and streaming setting. We give the first efficient deterministic distributed and streaming algorithms for l_p -regression for all $p \geq 1$, including infinity, entrywise l_p -low rank approximation, and versions of approximate matrix multiplication. We also empirically validate our approach and show that it can be beneficial in practice.

Differential Privacy using Bisimulation

David Purser · Presentation at 1:50pm

Differential privacy ensures that the output of a program on two databases that differ by just one record will produce an output that is basically the same, ensuring the privacy of an individual is not compromised by being in the data or not. Bisimulation checks that two configurations within a formal system behave equivalently, or in the probabilistic case measure how much they differ. Bisimulation has the possibility to verify that differential privacy holds in a formal system. The talk will explore how this is the case and the current research into the verification of bisimulations and current research in the automated verification of differential privacy.

Privacy Preserving Data Aggregation And Marginal Release

Tejas Kulkarni · Presentation at 2:05pm

The model of differential privacy has emerged as an accepted model to release sensitive information while giving a statistical guarantee for privacy. We focus on two problems.

1) For the problem of count queries, we design mechanisms to release the count associated with a group of n individuals. Prior work has focused on designing mechanisms by raw optimization of a loss function, without regard of integrity constraints on the output. This provides mechanisms which blindly optimize an error metric but which lack interpretability. We address these concerns by introducing a set of desirable properties that mechanisms can obey.

2) Imagine a survey in which participants answer d sensitive binary questions. The untrusted survey conductor is not allowed to aggregate each user's response in clear. This means each user has to add noise to her data locally. The conductor is interested in finding associations in the distribution of $k \leq d$ questions for the entire population.

Compilers: the Practical Benefits of a Theoretical Approach

Eleanor Davies · Presentation at 2:20pm

Compilers are integral to the execution and success of many applications. The complexity of realistic compilers means that tasks such as debugging them can be difficult and time consuming. Compiler verification eliminates these bugs and delivers guarantees where experimental results cannot. In mission or safety critical situations, unexpected behaviour can have disastrous consequences. Therefore the certainty provided by verification is particularly valuable. Using interactive theorem provers, such as Coq, strengthens these assurances by producing machine checked correctness proofs. Coq also provides some degree of automation which can simplify the complex verification process.

This work looks at existing theoretical approaches to compiler verification, their practical implications, and how they may be exploited. It also demonstrates the process of verifying compiler correctness using a simple example in Coq. Finally it identifies some possibilities for applying these methods to a language such as Scala, which is used in many large scale applications.

Resource and Makespan Minimization for Time-sharing DAG Application

Shenyuan Ren · Presentation at 2:35pm

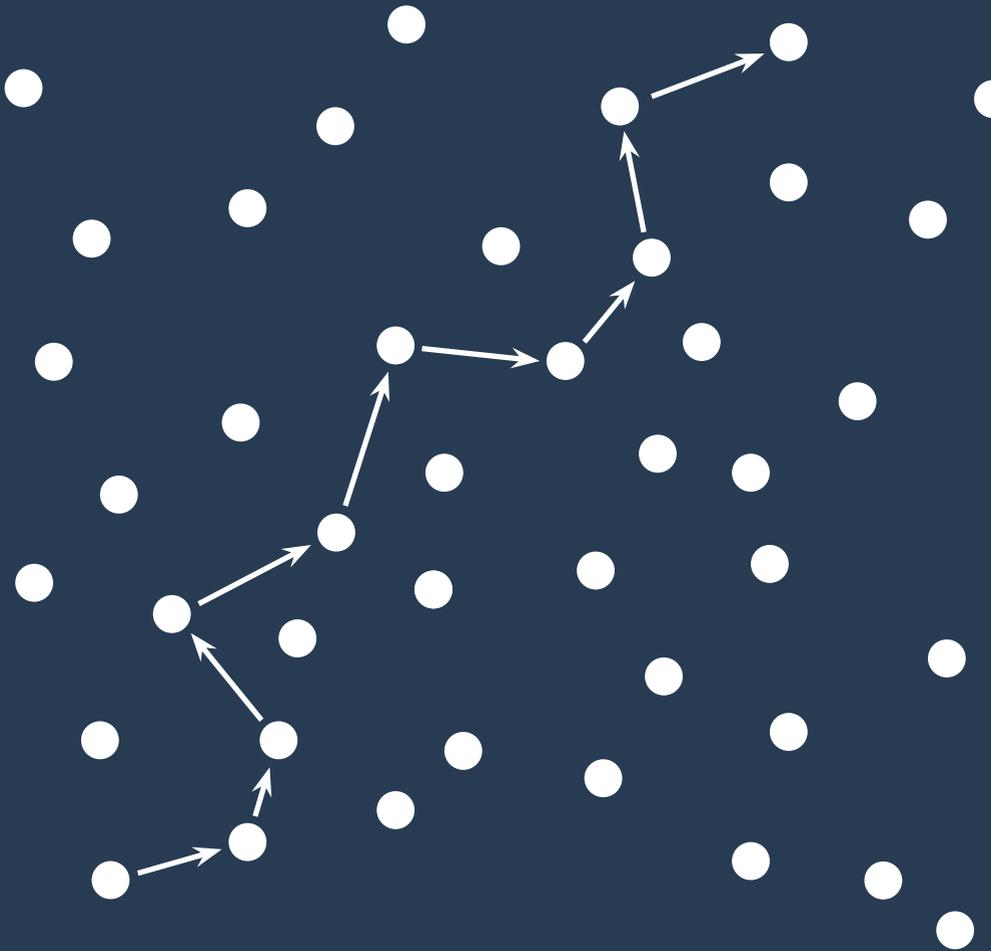
The essence behind task allocation and DAG makespan minimization problems is proven to be NP-complete when there are more than two resources (i.e. processors). Thus many task allocation and schedule strategy approaches are developed to minimize the DAG makespan and improve execution efficiency, and makespan models are developed to calculate DAG makespan. Tasks are regarded as executing one by one within a processor in these traditional common makespan models. In fact, when allocated on a specific processor, tasks are scheduled by the operating system scheduler in a round-robin manner consuming time slots. Tasks don't need to wait for its previous task (in the queue) to be finished before starts. Instead, tasks will execute in a time-sharing way along with all other runnable tasks within a processor. It is obvious that task finish time under time-sharing condition is different from that under non-time-sharing (i.e. one by one) condition. To our best knowledge, we are the first to develop a makespan calculation model taking time-sharing scheduling into account. Our time-sharing makespan model works out the task finish time among more than one concurrently executed tasks, thus works out the real total makespan of a DAG. Given a Schedule (DAG, Resource, Mapping), we set the expected makespan calculated by traditional non-time-sharing makespan model as a deadline. In our work, a task allocation adjustment strategy is developed to try out best to reduce to real makespan to corresponding deadline. Sometimes the real makespan can't be reduced to deadline by only adjusting. Thus a task allocation strategy is developed to reallocate the DAG tasks to the resources (given in the Schedule). If the deadline requirement still can't be satisfied, we will increase the processor number to further reduce the real makespan. Thus a task allocation with uncertain processor number strategy is developed so that minimum number of processors are used to meet the deadline requirement. In addition, bounds of the processor number are formulated to further improve efficiency.

Parallelizing the All-Pairs Shortest-Path Algorithm for Large Scale Graphs

Mohammed Alghamdi · Poster presented in foyer

We present a new method for solving the All-Pairs Shortest-Path (APSP) problem for big graphs by partitioning and processing it in parallel. In this work Breath First Search Algorithm is used to partition the graph and the Floyd-Warshall algorithm or Dijkstra's algorithm to find the APSP in each part. We implemented two ways to combine the output of each part of the partitioned graph and find the APSP for the whole graph.

Networking



Introduction by Dr. Arshad Jhumka

Chaired by Richard Kirk

Monitoring Recurrence of Salmonella Typhi in Outpatients

Betty Agbons · Presentation at 1:35pm

This paper describes TyfeSensor, a wearable, wireless Body Sensor Networks for continuous monitor of vital signs in Typhoid Fever outpatient. In contrast to previous systems where checkup is only possible when patient makes sometime long journey back to the hospital or to a trained health worker. This is a short-team solution. Sometimes patient are unable to make the journey back to the health Centre, particularly those living in rural areas due to mobility, resource constraints, lack of effective communication network, logistic limitations. In contrast TyfeSensor is designed to support long-term, longitudinal monitor and data collection on patient that have suffered from Typhoid Fever. Patients wear a sensor device and download the TyfeSensor APP for android and iPhone to monitor accurate body temperature and other physiological conditions any time and place. TyfeSensor is designed to overcome the core challenges of power supply, long battery duration to support the number of hours patient wears the sensor device per day and level of energy consumption in order to achieve true pervasive health monitoring for Typhoid Fever out patient.

Automating Cloud SLA Violation Prevention using Fair Exchange Protocol

Farrukh Qazi · Presentation at 1:50pm

Managing an “exchange fairness” is extremely paramount, within the modern e-commerce realm. Evaluating the extent of fairness varies from tangible items over the non-tangible items like digital data, quality of service (QoS) etc. The Cloud Service Providers (CSP) provision multi-flavored on-demand services to Cloud Service Subscribers (CSS). Service Level Agreements (SLAs) are signed by CSS and CSP, as a legal instrument, stating service definitions and monetary commitments. A malevolent CSP could misbehave by compromising it's services. It also shifts the entire burden of proof to the victimized CSS, for evidencing the SLA violation. This presents a serious security and a financial ambiguity to the CSS. A dishonest party e.g. CSP eventually dispute legit service credits claimed by the CSS. To patch this ultimate vulnerability, we introduce an architecture based upon Fair Exchange Protocol (FEP), which automatically advocates exchange fairness and dispute resolutions. The architecture administers least involvement of a Trusted Third Party (TTP), who intelligently takes controls to sort a service dispute.

A Near-Optimal Source Location Privacy Scheme for Wireless Sensor Networks

Matthew Bradbury · Presentation at 2:05pm

As interest in using Wireless Sensor Networks (WSNs) for deployments in scenarios such as asset monitoring increases, the need to consider security and privacy issues also becomes greater. One such issue is that of Source Location Privacy (SLP) where the location of a source in the network needs to be kept secret from a malicious attacker. Many techniques have been proposed to provide SLP against an eavesdropping attacker. Most techniques work by first developing an algorithm followed by extensive performance validation. Differently, in this presentation, the SLP problem is modelled an Integer Linear Programming optimization problem. Using the IBM ILOG CPLEX optimiser, we obtain an optimal solution to provide SLP. However, that solution is centralised (i.e., requires network-wide knowledge) making the solution unsuitable for WSNs. Therefore, we develop a distributed version of the solution and evaluate the level of privacy provided by the distributed solution. The solution is hybrid in nature, in that it uses both spatial and temporal redundancy to provide SLP. Results from extensive simulations using the TOSSIM WSN simulator indicate a 1% capture ratio is achievable as a trade-off for an increase in the delivery latency.

On the Generation of TDMA Schedules in Source Location Privacy-Aware Wireless Sensor Networks

Jack Kirton · Presentation at 2:20pm

Source location privacy (SLP) is becoming an increasingly important property for wireless sensor networks used in asset monitoring, where the source of messages is kept hidden from an eavesdropping attacker. Most works on SLP have focused on perturbing the message routing protocol to either draw the attacker away from the source or delay the attacker reaching the source. In this work, we conjecture that similar traffic perturbation can be achieved at the link layer level through suitable assignment of time slots to nodes. Thus, we focus on the problem of generating SLP-aware TDMA schedules for data aggregation scheduling. As the general optimisation problem for TDMA schedules is NP-complete, we propose the use of genetic algorithms to generate such schedules. We show that the solutions produced have a near optimal capture ratio and a high packet delivery ratio.

Exploiting Spontaneous Transmissions for Broadcasting and Leader Election in Radio Networks

Peter Davies · Presentation at 2:35pm

In the field of distributed computing, the radio network model is a classical and well-studied model of wireless communication networks. We study two fundamental tasks in the model: broadcasting and leader election. By making use of spontaneous transmissions (transmissions made before receiving the broadcast message) we design a protocol which improves the asymptotic running time for both of these tasks over the previous best results. Furthermore, our leader election time matches our broadcasting time, which is surprising since leader election is usually a harder task. This talk is based on joint work with Artur Czumaj.

Monitoring and Debugging in Wireless Sensor Networks - Survey

Al Alharbi · Poster presented in foyer

WSN applications are facing problem from unknown bugs and frequent failures due to their characteristics , distributed architecture, concurrent execution and resource limitations State how presentation will benefit audience. The aim of a WSN debugger is to iteratively detect and find the root causes of the failures which are difficult to identify with limited resources and visibility of the WSN network. It is not easy to achieve active debugging on WSN applications,because most of the applications are context sensitive and event driven. It is usually cannot fully control their operating context and triggering events. Therefore, there are many debugging techniques have been proposed and developed from different researchers and developers which explained in details in this paper. This paper is a survey conducted on the different implementation techniques for WSN debuggers. WSN debugging techniques can be classified based on their usage in application life or based on implementation strategies.

Towards Selecting Routing Protocols for Source Location Privacy Aware Sensor Networks

Chen Gu · Poster presented in foyer

Source location privacy (SLP) is becoming an important property for a large class of security-critical wireless sensor network applications such as monitoring and tracking. Experiments have been conducted to that the proposed protocols could enhance the level of SLP imparted to the network, under various attacker models and other conditions. Since there is a plethora of protocols, some of those with slight variations or configurations that make them more suitable to particular situations, it is difficult to select a SLP protocol that will provide the best trade-offs across a set of possibly conflicting attributes. In this paper, we propose a methodology where protocols are first profiled to capture their performance under various protocol configurations. Then, we present a novel decision-theoretic procedure for selecting the most appropriate SLP routing algorithm for the application and network under investigation. We show the viability of our approach through different case studies.

Cross-Color Channel Perceptually Adaptive Quantization for HEVC

Lee Prangnell · Poster presented in foyer

HEVC includes a Coding Unit (CU) level luminance-based perceptual quantization technique known as AdaptiveQP. AdaptiveQP perceptually adjusts the Quantization Parameter (QP) at the CU level based on the spatial activity of raw input video data in a luma Coding Block (CB). In this paper, we propose a novel cross-color channel adaptive quantization scheme which perceptually adjusts the CU level QP according to the spatial activity of raw input video data in the constituent luma and chroma CBs; i.e., the combined spatial activity across all three color channels (the Y, Cb and Cr channels). Our technique is evaluated in HM 16 with 4:4:4, 4:2:2 and 4:2:0 YCbCr JCT-VC test sequences. Both subjective and objective visual quality evaluations are undertaken during which we compare our method with AdaptiveQP. Our technique achieves considerable coding efficiency improvements, with maximum BD-Rate reductions of 15.9% (Y), 13.1% (Cr) and 16.1% (Cb) in addition to a maximum decoding time reduction of 11.0%.

Proactive Database Forensics: Chain of Custody Requirements for Database Audit Records

Denys Flores Armas · Poster presented in foyer

In forensic database investigations, audit records are important evidence for analysing malicious activities carried out by trusted employees or insiders who may have misused their privileged access to sensitive transactional information. Our research proposes a proactive forensic approach for the generation, collection and preservation of audit records related to insider activity in order to ensure accountability whilst satisfying Chain of Custody (CoC) requirements. Hence, role segregation, evidence provenance, event timeliness and causality are considered as functional requirements for the implementation of a forensically ready architecture to proactively investigate insider activity. This architecture implements triggers and stored procedures as forensic routines for building a vector-clock-based timeline in order to explain causality in suspicious transactional events that can be attributed to malicious insiders.

Urban Analytics



Introduction by Prof. Graham Cormode

Chaired by Liam Steadman

Understanding Happiness in Cities Using Social Media

Neha Gupta · Presentation at 1:35pm

The demographics and landscape of cities is changing rapidly, and there is an emphasis to better understand the factors which influence citizen happiness. Few research outputs have attempted to understand how large-scale sentiment data maps to urban socioeconomic and infrastructure features. Inferring happiness (sentiment analysis) from social media data is a scalable solution. In this paper, we apply natural language processing to 0.4 million geo-tagged Tweets in the Greater London area to understand the underlying socioeconomic and urban geography parameters that influence happiness. Our results not only verify established thinking: that job opportunities correlate strongly with positive sentiments; but also reveal two additional insights: (1) happiness has a negative relationship with the number of children and (2) happiness has a parabolic relationship with access to public transportation. The latter implies that those that rely on, but do not have strong access to public transport are the least happy.

Using the Weather Research and Forecasting model to validate and improve existing Unified Model mesoscale calculations

Timothy Sit · Presentation at 1:50pm

When resolution is refined to sub-kilometre scales, the non-hydrostatic mesoscale version of the Unified Model (UM) does not converge to what is observed in and around convective complexes. However the Weather Research and Forecasting (WRF) model can take advantage of its ability to nest finer grids within coarser meshes to simulate conditions as fine as 100 metres, making the calculations relevant to urban environments. My plans will be to use WRF to compare against existing UM mesoscale calculations that have themselves been compared with field data. The goal will be to provide a set of reliable high-resolution reference simulations over a range of environmental conditions that can be used to understand why the UM calculations are not agreeing with the observations, and to improve parameterisations in the UM to resolve these problems.

Locating food deserts and predicting diet-related health risks with social media

Elisabeth Titis · Presentation at 2:05pm

Motivated by knowledge gaps in the evidence base (Wrigley 2002), this research will utilise metadata in social media as social ‘sensors’ to investigate links between dietary patterns and nutrition-related ill-health, focusing on low-income and social exclusion, and poor accessibility in terms of retail geography to inform public health policy and estate development. The novelty is threefold: a) the combination of conventional datasets with new forms of data harvested from Twitter and Instagram to assess the adoption of social media platforms in terms of content and language usage as a way to gather better empirical evidence on the health inequalities agendas; b) mapping food deserts using spatial methods to visualise areas with the highest demand for improvement in healthy food access; c) studying links between food and health outcomes in relation to dietary acid load as a complementary way to study the aforementioned in relation to nutrition.

Spatial analyses of social media data to mitigate disease spread

Alex Caton · Presentation at 2:20pm

From the literature review performed in this study, it is apparent that very little research tackles the use of spatial methods in the study of social media and disease tracking, and that future research should be dedicated to rectifying this. This study highlights the areas that are currently undertaking this research. In addition, this study investigates existing literature on how diseases are tracked, finding correlation between social media and developing methodologies. Working with existing research, this project finds current trends in digital epidemiology, informing researchers of gaps and potential avenues for future study. Furthermore, using both semantic and spatiotemporal information from location-based social networks such as Twitter can provide new and innovative ways to study epidemics. This presentation aims to perform a review of the current research on the spatial methods used with social media data for tracking and/or predicting the spread of infectious diseases.

Predicting floods with Flickr tags

Nataliya Tkachenko · Poster presented in foyer

Increasingly, time and location stamps of social media postings and other types of user generated content (UGC) are being used to provide useful operational information during such natural hazard events as hurricanes, storms and floods. As choices of semantic tags in current methods are usually reduced to the exact name or type of the event (e.g., hashtags ‘#Sandy’ or ‘#flooding’), the main limitation of such approaches remains their mere nowcasting capacity. In this study we make use of polysemous tags of images posted during several recent flood events and demonstrate how such volunteered geographic data can be used to provide early warning of an event before its outbreak.

A Spatio Temporal, Gaussian Process Regression, House Price Predictor

Henry Crosby · Poster presented in foyer

The primary aim of real-estate value prediction differs significantly from one stakeholder to another: (1) A developer is looking to maximise their returns; (2) Lenders seek to minimise their risk and hence are interested primarily in the market form efficiency of real estate; (3) Home buyers have non-return related priorities, such as life-style suitability and location. Varying motives, macro-environmental factors and data sparsity are just some of the reasons why real estate valuation remains a challenging task. Additionally, unprecedented leverage (for example 100% loans) and market potential (a 4.5% increase in residential property prices in the UK in 2015) are some of the reasons why real-estate value prediction can be so valuable. This paper introduces a novel four-stage methodology for real-estate valuation. The spatio-temporal Gaussian process regression is trained on a sample of 16,000 estate transactions and is validated against regression kriging, random forests and an M5P-decision-tree with 231,000 instances utilising R2 and RMSE. The trained model is integrated into a real estate decision engine for commercial use. Model validation demonstrates a 96.6% accuracy.

Combining Heterogeneous User Generated Data to Sense Well-being

Adam Tsakalidis · Poster presented in foyer

We address a new problem of predicting affect and well-being scales in a real-world setting of heterogeneous, longitudinal and non-synchronous textual as well as non-linguistic data that can be harvested from online media and mobile phones. We describe the method for collecting the heterogeneous longitudinal data, how features are extracted to address missing information and differences in temporal alignment, and how the latter are combined to yield promising predictions of affect and well-being on the basis of widely used psychological scales. We achieve a coefficient of determination of 0.71 – 0.76 and a ρ of 0.68 – 0.87 which is higher than the state-of-the art in equivalent multi-modal tasks for affect.

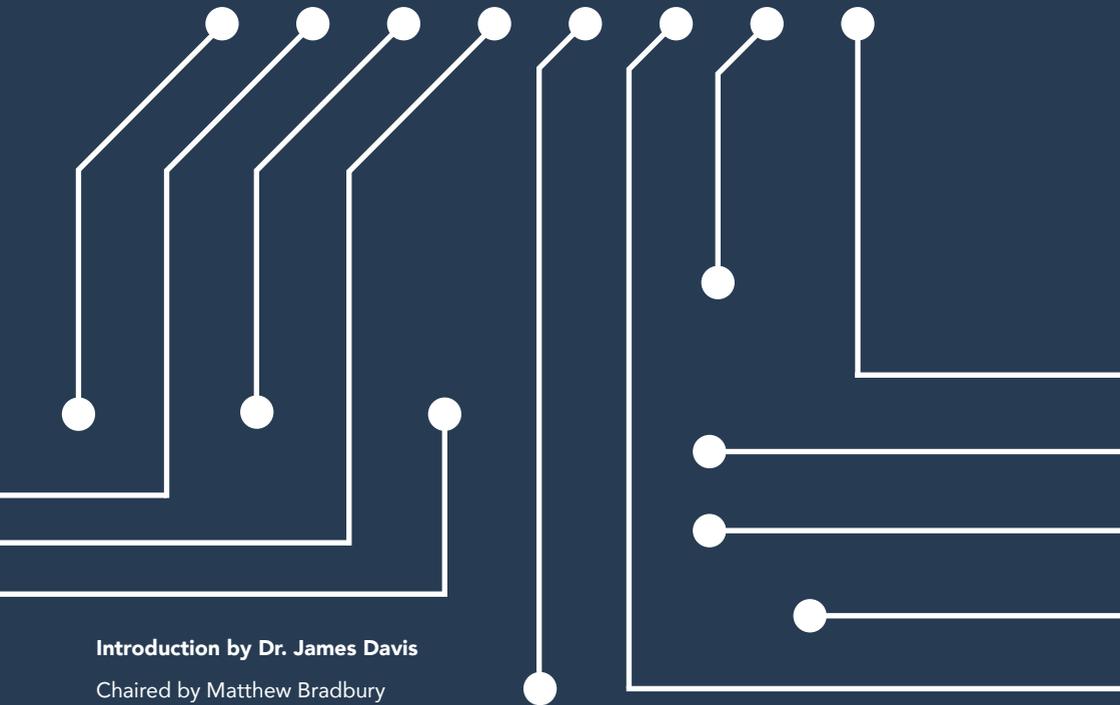
Identifying data trends to support current and future service planning in Birmingham

Sarunkorn Chotvijit & Malkiat Thiarai · Poster presented in foyer

There is huge national interest in tackling issues surrounding the needs of vulnerable children and adults. Child deaths that have made national headlines have highlighted the major challenges facing cities in protecting and safeguarding its most vulnerable citizens. The performance of a local authority in meeting these needs is the focus of the national regulator, Ofsted. This research is rooted in the outcomes of Ofsted inspections in Birmingham. In 2010, 2012 and 2014, Ofsted judged the provision of the services to children in Birmingham as inadequate, meaning that there are widespread, serious failures that leave children and young people at risk of harm. Our research involves extracting and analysing Birmingham City Council's social care data to tackle persistent failures in the service, which has, in turn, led to poorer outcomes for children and young people in the city.

TRACK 7 · OC1.07 · 3:40pm

High Performance Computing



Introduction by Dr. James Davis

Chaired by Matthew Bradbury

A New Approach to Neural Cryptography - Utilizing One-Time Pad in an Audio Cryptosystem

Jack Jackson · Presentation at 3:45pm

This research is conducted with the goal of exploring the potential of artificial neural networks to ensure key secrecy, authenticate and validate member parties, and sufficiently handle the attempted de-synchronization of symmetric key generation. We propose a method of recovering and reconfiguring member networks in such a way that the impact on system performance is negligible. The keys generated shall be applied to live audio data via a one-time pad method of encryption, which mathematically, is the only truly secure method of encrypting data. We provide an efficient alternative to modern audio cryptosystems, which is employable on the most modest of hardware, making it highly applicable to the majority of real world systems. Ultimately we aim to propose a method of implementing one-time pad encryption in combination with neural network key generation; ensuring that its key secrecy and reuse deficiencies are managed accordingly. We also present a method for the authentication and validation of each communication between valid member networks. We offer an alternative to the widely implemented tree parity machine, which offers potentially superior security whilst maintaining a similar degree of convenience regarding synchronization.

Predicting scaling performance of a production CFD solver

Andrew Owenson · Presentation at 4:00pm

Large and high-fidelity computational simulations are an essential component of modern scientific and engineering research, but their growing computational demand means considerable energy and computing facility access costs. Thus, it is important to ensure that the parallel performance of the simulation code is well optimised, and to understand the cost-benefit relationship of allocating more computing resource to a simulation. A predictive performance model of the code is an important tool for achieving these aims.

This works focuses on HYDRA, a computational fluid dynamics simulation solver in use by Rolls-Royce. An analytical performance model has previously identified a communication bottleneck which had reduced scaling performance. This work extends the performance model with stochastic terms that capture inter-process performance variation resulting from memory contention. This extension improves the accuracy of scaling performance prediction from limited benchmark data, enabling more informed decisions regarding computing resource allocation.

Novel Data Structures and Parallelisations using Mini-Applications

Richard Kirk · Presentation at 4:15pm

For many years, achieving higher performance of a program meant increasing the clock speed of the processor. However, this could not be maintained due to the increase in energy consumption. Thus, parallelism and multi-core systems had to be developed to continue increasing performance. Now, we have accelerator cards such as GPU's, as well as more many-core systems such as Xeon Phi. However, a lot of the legacy code was not written to exploit these high levels of parallelism seen in modern day High Performance Computing systems. As well as this, due to large range of different systems, making sure these legacy code are “future proof” and performance portable can be very difficult. This difficulty is only amplified when the code has a complex data structure and access pattern. My aim is to investigate how to parallelise these data structures efficiently, making them perform better with multi-core systems.

Higher Order Algorithms for Particle-In-Cell Simulations

Dom Brown · Presentation at 4:30pm

The particle in cell (PIC) method is widely used within physics codes in order to simulate the behaviour of charged particles within plasmas. In unstructured PIC codes, basis functions are employed in order to facilitate the process of interpolating the electric and magnetic fields from grid points to particle locations, and also when particles deposit charge back onto the grid. Using second order basis functions allows for more accurate solutions to be achieved versus a first order function at the same coarseness of grid, but this comes at a greater computational cost.

This presentation examines the implementation of higher order methods within mini-PIC, a PIC code developed by Sandia National Laboratories, and aims to investigate the performance of the code on Graphics Processing Units (GPUs).

All-Reduce for Distributed Machine Learning

Zhenyu Li · Presentation at 4:45pm

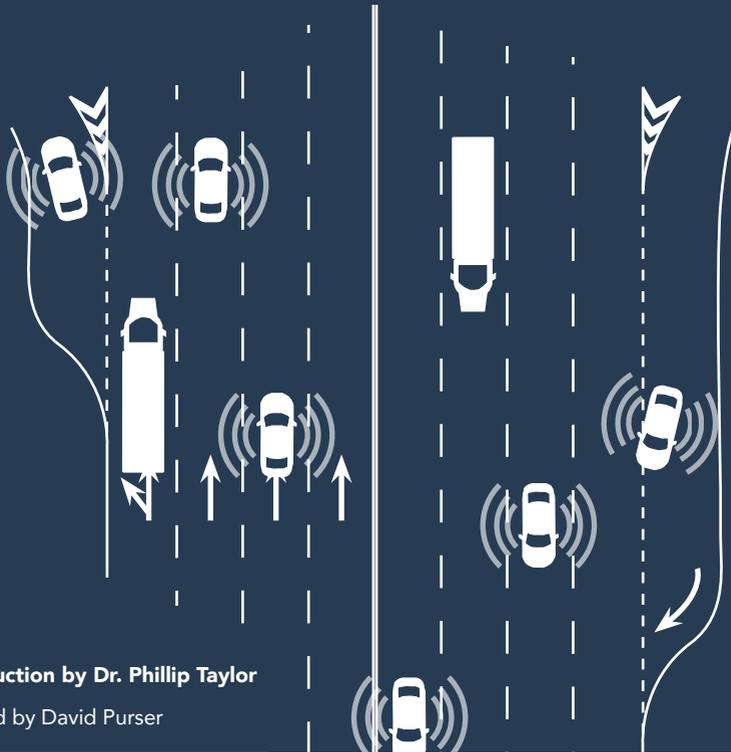
All-Reduce is a collective operation that combines elements in the given input vector by an associative and commutative function, and the result is received by the participating processes. It is a well-known operation in the message-passing interface (MPI) for high-performance scientific computing. In machine learning, deep learning in particular, synchronous weight-updates in the distributed Stochastic Gradient Descent (SGD) relies on all-reduce. However, current implementations are based on the centralized parameter server approach that creates bottleneck at the root process, which is not only inefficient but also underutilizing the resources in a high-performance cluster. We present an efficient all-reduce implementation on Apache Spark, which is a mainstream data-analytic framework. We report a 3-9x speed-up against the original reduce+broadcast approach in Spark.

Performance Portability of a Representative Unstructured ALE mini-app

David Truby · Poster presented in foyer

This poster concerns an unstructured 2D mini-app that uses Arbitrary Lagrangian-Eulerian methods. The mini-app is designed to evaluate the utility of various architectures for unstructured mesh applications, an area that has seen much less investigation than structured applications, without requiring a large scientific application to be run. Implementations of the app exist for various programming models covering both GPUs and CPUs, allowing the performance of the application on various architectures and using various programming models to be evaluated. In particular the main models to be evaluated are the MPI, OpenMP, OpenACC and CUDA models, and testing has been done on 22-core Broadwell CPUs, Intel Xeon Phi Knights Landing CPUs and Nvidia K20X GPUs. Current results show that the best performance on this application is given using traditional CPUs and programming models, although code refactoring and redesign may allow better performance to be obtained when using accelerators.

Artificial Intelligence & Transport



Introduction by Dr. Phillip Taylor

Chaired by David Purser

Transfer Learning for Concept Drifting Data Streams

Helen McKay · Presentation at 3:45pm

Bespoke applications provide personalised functionalities to users but can be expensive to produce. To reduce costs, off-the-shelf applications utilise manual input to tailor functionalities to user's personal preferences. Manual inputs must be frequently updated and maintained to ensure they effectively represent the user's desires. Online machine learning techniques can be used to mitigate manual inputs, predicting personalised preferences from sensing the surrounding environment. User preferences and real-world environments are dynamic in nature, causing concept drifts that require learning algorithms to update or change their predictive models to remain effective. To build effective models, learning algorithms require initial training data, preventing useful predictions from initially being made for new users, this is known as the cold start problem. Transfer learning can be utilised to overcome the cold start problem by transferring knowledge, in the form of learnt models, from existing users to new users. This enables existing models to be utilised before sufficient data is collected to learn the current concept. A novel framework has been developed to transfer knowledge between two online data streams with the overarching aim of improving the effectiveness of real-time predictions.

Vehicle Visit Detection using In-car Data

James Van Hinsbergh · Presentation at 4:00pm

GPS trajectories can be obtained with reasonable accuracy using standard off-the-shelf hardware. Using Gradient-based Visit Extraction, these trajectories can be used to calculate visits of the users. In this work, we apply this technique in the context of vehicles, highlighting locations in which the user has been. Given initial parameter optimisation, the problem of false visits occurs, notably in heavy traffic where the vehicle is stationary for a considerable period of time. To isolate these visits, we build a binary classifier using in-car data to determine whether or not the visit is of interest. The classification model uses Random Forest with Principle Component Analysis to select the in-car features to use, achieving high accuracies.

Convention Emergence in Partially Observable Topologies

James Marchant · Presentation at 4:15pm

In multi-agent systems it is often desirable for agents to adhere to standards of behaviour that minimise clashes and wasting of (limited) resources. In situations where it is not possible or desirable to dictate these standards globally or via centralised control, convention emergence offers a lightweight and rapid alternative. Placing fixed strategy agents within a population, whose interactions are constrained by an underlying network, has been shown to facilitate faster convention emergence with some degree of control. Placing these fixed strategy agents at topologically influential locations (such as high-degree nodes) increases their effectiveness. However, finding such influential locations often assumes that the whole network is visible or that it is feasible to inspect the whole network in a computationally practical time, a fact not guaranteed in many real-world scenarios. We present an algorithm, PO-Place, that finds influential nodes given a finite number of network observations. We show that PO-Place finds sets of nodes with similar reach and influence to the set of high-degree nodes and we then compare the performance of PO-Place to degree placement for convention emergence in several real-world topologies.

Limiting Concept Spread in Environments with Interacting Concepts

James Archbold · Presentation at 4:30pm

The propagation of concepts in a population of agents is a form of influence spread, which can be modelled as a cascade from an initial set of individuals. In real-world environments there may be many concepts spreading and interacting. Previous work does not consider utilising concept interactions to limit the spread of a concept. In this paper we present a method for limiting concept spread, in environments where concepts interact and do not block others from spreading. We define a model that allows for the interactions between any number of concepts to be represented and, using this model, develop a solution to the influence limitation problem, which aims to minimise the spread of a target concept through the use of a secondary inhibiting concept. We present a heuristic, called maximum probable gain, and compare its performance to established heuristics for manipulating influence spread in both simulated small-world networks and real-world networks.

Methods for linking and analysing disparate datasets

Liam Steadman · Presentation at 4:45pm

Linking data is the process of correlating, associating and combining information from one or more sources into a single output. Linking and analysing disparate datasets is an area of research that spans multiple disciplines and touches almost all industries. The process is not concerned necessarily with concatenating one dataset onto another, it is concerned with the extraction of relevant information that is only possible with data from multiple sources. The talk will introduce several methods previously employed for working with disparate datasets and discuss future work in the field. Methods will be discussed for linking and analysing geospatial and temporal data, streaming data, trajectory data and graph data.

Bootstrapping Trust and Stereotypes with Tags

Caroline Player · Poster presented in foyer

Agents joining multi-agent systems (MAS) face two significant problems: they do not know who to trust and others do not know if they are trustworthy. Our contribution extends trust and stereotype approaches to use a comparison of agents' observable features, called tags, as an initial indication of expected behaviour. The results show an improvement in agents' rewards in the early stages of their lifetimes, prior to having sufficient information to use trust or stereotype methods.

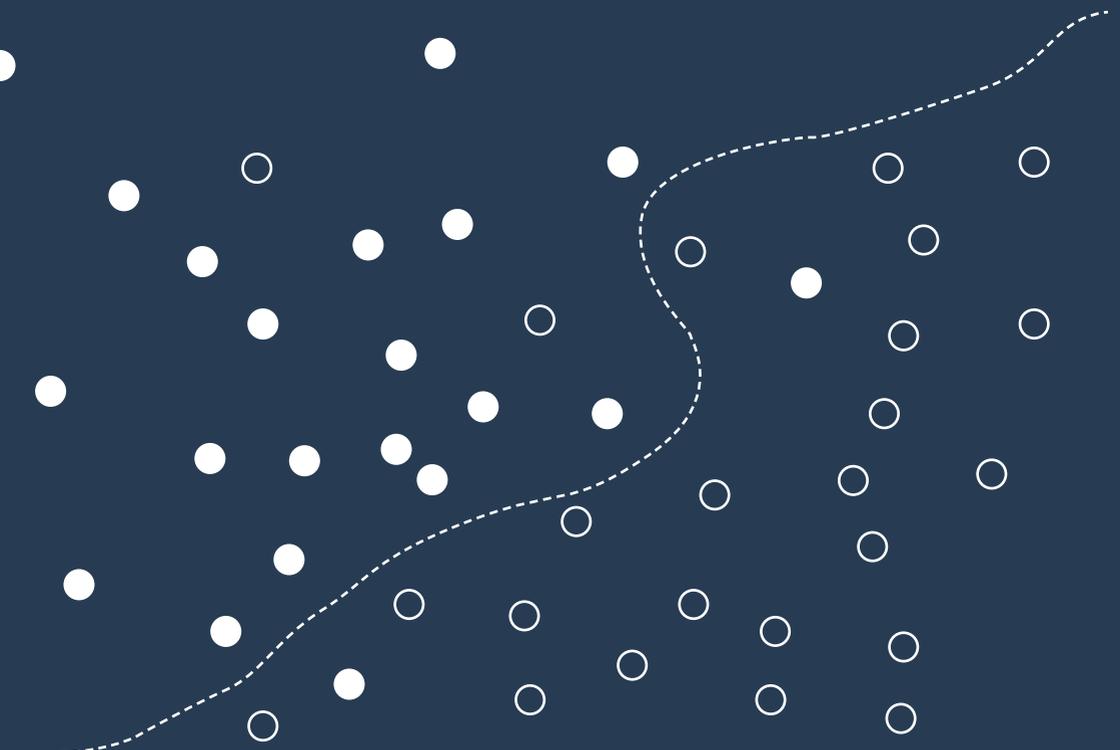
Concept change in machine learning

Adam Gelencser · Poster presented in foyer

The volume of data generated and captured within automotive industry during all life cycle stages is exponentially growing. There are multiple initiatives how the available data can be utilised to increase reliability, durability, efficiency and driver comfort. This is the case within Jaguar Land Rover as well, where data mining and machine learning tools are more widely used than ever before. This research considers concept drift detection in the context of vehicle data. Concept drift occurs within dynamically changing environments where the underlying distribution of inputs or desired outputs can change significantly and to ensure accurate predictions, the models learnt must be updated accordingly to the rate of concept drift.

TRACK 9 · OC1.09 · 3:40pm

Machine Learning



Introduction by Dr. Theo Damoulas

Chaired by Greg Watson

Accelerating the Training Process of a Deep Neural Network via improving the Weight Updating Algorithm

Junyu Li · Presentation at 3:45pm

Deep Learning is currently a greatly active research area in the machine learning and pattern recognition area. When scaling up Deep Learning algorithms, it has been shown to lead to increased performance in benchmark tasks and to enable the discovery of sophisticated high-level features; on the other hand, it also presents enormous challenges where at the core is the training efficiency of the Deep Neural Network. Nearly all of the Deep Learning is powered by one very important training algorithm: Stochastic Gradient Descent (SGD). However, there exists a heavy data dependence in SGD training which extremely limits the degree of parallelism. This research purposes to seek for an improvement to SGD to speed up training process of the network. The current work tries to combine regular SGD with parallelism enabled algorithms, such as Mini-Batch SGD, which allows for parallel processing within each batch of data, so that SGD can take the benefits of parallel processing.

Mining similar subsequences in time series efficiently

Zhuoer Gu · Presentation at 4:00pm

Time series data are ubiquitous in real life. Many time series data mining and analysing tasks require knowledge of time series similar subsequences, aka motifs, suggesting the fundamental importance of mining motifs in time series data. However, constrained by the natural difficulty of motif mining, current motif mining methods are either computationally expensive or unable to perform accurate mining. In this paper we propose a novel method to achieve fast and accurate time series motif mining. The method first locate a small set of possible start positions of motifs quickly, then search locally for the best matched motifs. The method has outstanding computational and spatial efficiency. Experiment results show the method's capability of discovering time series motifs with low requirement of computational resource.

Verification of Outsourced Data Analysis

Christopher Hickey · Presentation at 4:15pm

As the popularity of outsourced computations increases, concerns about accuracy and trust between the client and the cloud computing service become ever more relevant. Our work aims to provide faster and more practical methods to verify analysis of large data sets, where the client's memory costs are independent of the size of the data set. We do this by using annotated data streaming methods, in which the cloud computing service provides a short proof alongside the results which can be used to confirm the correctness of the computation. We supply an optimally efficient protocol for verifying matrix multiplication, and use this to provide protocols for ordinary least squares (OLS) and principal component analysis (PCA).

Enabling Data-Driven Dependability Analysis from Cluster Log Data

Edward Chuah · Presentation at 4:30pm

Recent work have used both failure logs and resource use data separately (and together) to detect system failure-inducing errors and to diagnose system failures. System failure occurs as a result of error propagation and the (unsuccessful) execution of error recovery mechanisms. Knowledge of error propagation patterns and unsuccessful error recovery is important for more accurate and detailed failure diagnosis, and knowledge of recovery protocols deployment is important for improving system reliability. In this talk, I will give an overview of recent work in failure diagnosis, error detection and failure prediction, present the CORRMENT framework which carries failure diagnosis another significant step forward by analyzing and reporting error propagation patterns and degrees of success and failure of error recovery protocols, and conclude with a discussion of the future work.

Twitter Analysis to Predict the Satisfaction of Telecom Company Customer

Latifah Almuqren · Presentation at 4:45pm

This research aims at mining Arabic tweets to measure customer satisfaction toward Telecom companies in Saudi Arabia and to predict the ratio of customer churn. The contribution of this study will be capitalised as recommendations to these companies, based on monitoring in real time their customers' satisfaction on Twitter. Some steps already started, such as build a corpus of Arabic SSA messages via the Twitter semantic (search) API, using a Python script searching for real time tweets that mention Telecom companies using the hashtags to monitor the latest sentiments of Telecom customers continuously. The subset is 20,000 tweets that are randomly selected from the dataset, for training the machine-classifiers SVM and NB. The results showed the SVM is achieved better accuracy than NB with Arabic text.

Sequential Approach to Rumour Stance Classification

Elena Kochkina · Poster presented in foyer

Rumour stance classification is a task of identifying the attitude of Twitter users towards the truthfulness of the rumour they are discussing as either Supporting, Denying, Questioning or Commenting. Rumours that attract a lot of skepticism are more likely to be proven false later (Zhao et al., 2015). We classify stance of a set of Twitter posts discussing rumours as part of SemEval 2017 Task 8 Subtask A RumourEval challenge. We propose a LSTM-based sequential model that models the linear structure of the conversation, and achieves an accuracy of 0.784 on the RumourEval test set, which outperforms all other systems in Subtask A.

TDParse: multi-target-specific sentiment recognition on Twitter

Bo Wang · Poster presented in foyer

Existing target-specific sentiment recognition methods consider only a single target per tweet, and have been shown to miss nearly half of the actual targets mentioned. We present a corpus of UK election tweets, with an average of 3.09 entities per tweet and more than one type of sentiment in half of the tweets. This requires a method for multi-target specific sentiment recognition, which we develop by using the context around a target as well as syntactic dependencies involving the target. We present results of our method on both a benchmark corpus of single targets and the multi-target election corpus, showing state-of-the-art performance in both corpora and outperforming previous approaches to multi-target sentiment task as well as deep learning models for single-target sentiment.

Using Recurrent Neural Networks (RNNs) to Improve Accuracy in Automatic Music Transcription

Martina Kluvancova · Poster presented in foyer

Automatic Music Transcription is the process of converting some form of musical audio (such as a .wav file) into some form of musical notation (such as a MIDI file or a piano roll representation). We propose a method that could be used to correct mistakes in transcriptions (and thus improve accuracies) by using musical knowledge. Music contains many different patterns and follows (and breaks) rules. This is why humans have expectations about what might happen next in a piece of music. We train a Long Short Term Memory Recurrent Neural Network (LSTM-RNN) on piano roll representations of music to learn this kind of musical knowledge. Then if we are uncertain about certain notes in a transcription, we can use the LSTM-RNN to predict most likely notes at that point in time given the musical context.

“Being a scientist means living on the borderline between your competence and your incompetence. If you always feel competent, you aren't doing your job.”

Carlos Bustamante

“People think that computer science is the art of geniuses but the actual reality is the opposite, just many people doing things that build on each other, like a wall of mini stones.”

Donald Knuth