Networks

Data Mining in Wales

Swansea University, 17-18 January 2006

We all want safer streets, and policing is a key issue. Rick Adderley, founder of AE Solutions, will speak on 'The use of Data Mining in Operational Policing'.

It is my great pleasure to invite NCAF to Swansea for a second time. We will be revisiting the popular theme of Data Mining and there has been an exceptional response. Almost all the papers are theme papers and experts from industry or academia will give the majority of them. I am really grateful to all the speakers for their time and support for this NCAF meeting.

The meeting will start with a presentation on 'Data Mining for Compliance and Audit' from Steve Woods, Technical Director at Informovers Ltd. The second presentation will address our social needs. We all want safer streets, and policing is a key issue. Rick Adderley, founder of AE Solutions, will speak on 'The use of Data Mining in Operational Policing'.

The analysis of large quantities of data requires a different approach, often involving Grid Computing. Professor Jim Austin, founder of Cybula Ltd, will give a presentation on 'High performance data pattern matching with AURA and Cortex II'. This will be followed by a presentation on 'Building better data mining models with ROC analysis' by Professor Peter Flach from Bristol University. Abhijeet Kashape, Vice President of Ontract Systems Ltd will be presenting a case study on 'Mobile communication in Data Mining'. Finally, Dr Paul Temblet from the NHS will talk on 'The doctor's perspective on Data Mining in the NHS'.

The Data Mining theme will continue on the second day with a presentation from the CERCIA group at Birmingham University. Professor John Tucker from Swansea University will speak on 'Data Centric Computation' giving it a theoretical computer science flavour. Dr Jyoti Sinha will share his industrial experience by presenting some case studies in signal processing. Professor Gyan Pande from Swansea University will talk on 'Mining experimental data for numerical computation'. Nazri Nawi, also from Swansea, will present his research on improving learning in back propagation neural networks by varying the adaptive gain.

I have not forgotten the social evening. It should be very interesting. A visit to the National Waterfront Museum that was opened to the public only a few weeks ago has been arranged. There will be plenty of food and fun! Allow us to extend our Welsh hospitality once again.

BURGLARS

POLICEMEN

Map from www.multimap.com

If only data mining were this easy
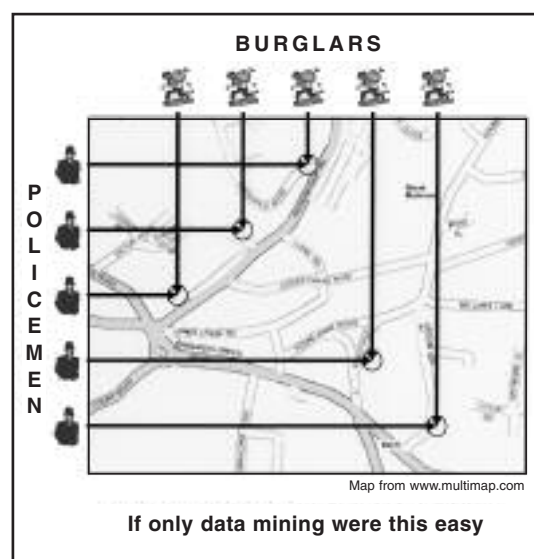
I am very much looking forward to seeing you all at Swansea in the New Year.

Dr Rajesh Ransing
School of Engineering
University of Wales, Swansea

It's that time of year again ...

As usual for our January meeting, the NCAF AGM will take place on the second day. This is for members only, and is a useful opportunity to review and comment on NCAF's performance and direction. The Committee currently has vacancies and members are encouraged to join it in order to be more pro-actively involved in the evolution of the organisation. Any new Committee members will be invited to introduce themselves at the AGM. The Committee meets no more than three times a year. The meetings are often arranged to coincide with the NCAF conferences.

Graham Hesketh, NCAF Chairman

Intelligent systems in sunny Southampton

The second day saw another excellent series of lectures and with the prospect of England saving the final Test to win the Ashes, there were feverish attempts to arrange wireless Internet access.

The autumn meeting of NCAF in Southampton saw fantastic weather, the prospect of the England cricket team winning the Ashes and an exciting programme of presentations on the theme of multi-agent systems.

The first day of presentations was held in the Nuffield Theatre on the University of Southampton campus, and despite a slightly recalcitrant data projector and an air conditioner that was not expecting to be called into service so late in the year, it all went off without a hitch. Chiu Fan Lee of Oxford University gave the first presentation of the meeting, and he described how probability collectives, a framework that ties together concepts from statistical physics, game theory, and distributed optimisation, could be applied to multi-agent systems. He described a setting, motivated by space exploration, whereby a group of robots are attempting to achieve a collective goal, but, due to communication restrictions, must choose their actions independently.

This presentation was followed by Rajdeep Dash, of the Intelligence, Agents, Multimedia Group at Southampton University, who described how game theory is being used in the design of future computational systems. Motivated by the actions of the prison guard in the famous Prisoner's Dilemma, he described how mechanism design allows the system designer to ensure that the overall system exhibits desirable global properties, despite the selfish actions of the individuals who use it.

The theme of multi-agent systems was continued by Daniel Kudenko of York University, who gave a presentation on how symbolic planning allows reinforcement learning to scale up, from toy problems, to more realistic real-world problems. The use of symbolic planning prevents the vast increase in the size of the state and action space that normally limits the applicability of these approaches.

Autonomous mailboxes

Next, Armin Stranjak of Rolls-Royce plc presented an elegant solution to the problem of reducing communication costs within a distributed messaging system. In his design, not only are the messages moving around the network, but the autonomous mailboxes themselves also move, in order to reduce traffic hot spots and delays.

This was followed by a more speculative presentation by Peter McBurney of Liverpool University, who introduced the notion of complex systems that are both anticipatory and reflective. This means that not only do the individuals within these systems model and attempt to predict the environment in which they find themselves, but they also talk to one another and share their predictions. Peter gave examples of these operating in the real world (including share traders and the Bank of England) and discussed some of the computational challenges that need to be met in order to build them using multi-agent systems.

Finally, rounding up the presentations for the day, Andrew Byde of HP Labs, described a real-world example of utility-computing that had been conducted with a group of animators in Bristol. In exchange for free access to HP computing resources, in order to render their animation frames, the animators were asked to manage a `budget' and `bid' against one another to purchase computer time. Andrew described the problems of taking these ideas out of the lab and actually trying to apply them in the real world. In particular, he discussed the problems of persuading non-specialist users to invest the time and energy required to become competent bidders within the market that he had constructed.

Having finished off the technical part of the programme, the day was not finished as it was now time for Puzzle Corner. This brought the unexpected pleasure of seeing the NCAF Chairman, Graham Hesketh, blindfolded in the corner of the lecture theatre. Andrew Byde who had solved the puzzle earlier, and had even prepared an interactive PowerPoint presentation, acted out the role of Professor Grey, and Fenella played Lisa. Needless to say, Lisa solved the problem very quickly

Having been entertained, a coach trip to the village of Warsash on the river Hamble followed. A boat then took the group on a tour of the river that passed through the middle of a fleet of dinghies taking part in the Warsash Sailing Club's evening regatta before heading upstream past the moorings of multi-million pound yachts. This was followed by a dinner at the 'Rising Sun' pub, which despite the name had a fantastic view of the setting sun.

Auction protocols

The second day saw another excellent series of lectures and with the prospect of England saving the final Test to win the Ashes, there were feverish attempts to arrange wireless Internet access. The first presentation was by Alex Rogers, of Southampton University and described how computer scientists are adopting auction protocols as a means of allocating resources and tasks within computer systems. In doing so, they are applying machine-learning techniques to allow the auctioneer to tailor the auction properties to the bidders who are participating.

This was followed by a talk that illustrated how even well established problems in computer science, such as graph colouring, continue to generate surprising results. John Shawe-Taylor, again from Southampton University, reviewed probabilistic approaches to this problem, and showed that recent insights into the structure of computationally hard problems can guide the use of novel message passing techniques such as survey propagation. Unlike conventional approaches that use a probabilistic description for the colour of each node, these approaches actually model the probability that a node's colour should be fixed (whilst still searching for the colours of the neighbouring nodes).

After coffee, Chris Satchwell of Technical Forecasts Ltd gave an overview of how pattern recognition techniques can be applied to trading decisions in the financial markets. Chris discussed the problems of applying fundamental analysis to stock prices including the difficulty of valuing a company that holds intangible assets such as brand awareness, patents and goodwill. As an alternative, Chris introduced the idea of technical analysis; specifically, developing measures, indicators and instruments that use historical stock prices to predict future price movements.

After lunch, Rishabh Gupta of Motorola described how they are using Bayesian nets to model and predict the quality of their software products. These models chart the introduction and removal of defects that occur throughout the software development lifecycle. By understanding where the defects are introduced, the model can be

The forensic analysis of print using digital image analysis

The forensic analysis of print is an important and active area of research in the fight against fraud and counterfeiting crimes. One of the objectives of the London College of Communication (LCC) research programme is to assist in this field. A major part of the forensic research programme at the LCC is to find better methods to 'fingerprint' printing machines.

A possible reason why it is often unfeasible to prosecute counterfeiters and document fraudsters is because it is very difficult to establish links between the counterfeiters or fraudsters and their printing equipment. This is due to the extremely wide range of high quality laser and inkjet printing machines available that can be bought cheaply. This problem is exacerbated by the fact that new varieties of these printing machines are being commercially produced on a continual and frequent basis.

Both laser and inkjet printers, which make up most of the cheap office printing market, have disposable ink and toner cartridges. This can make chemical analysis of ink and toner inconclusive since different toner and ink cartridges can have different chemical compositions. The chemical analysis is also a process that requires the destruction of part of the evidence.

The analysis of the actual shapes of the text characters produced by different print engines using a combination of digital image analysis and neural network techniques is another possible way of fingerprinting printing machines. However, this method not only suffers from the influence of replaceable ink and toner components distorting the results of the analysis, but other drawbacks as well. Firstly, processing huge amounts of image data due to the large number of makes and models of printing machines is time consuming. Secondly, the problem of ink spread on different kinds of paper or humidity conditions in inkjet printing processes distorts the shapes of text characters.

An identification methodology for fingerprinting printing machines recently considered is to use Electrostatic Detection Apparatus, or ESDA. It is hoped that ESDA will be able to aid existing digital image analysis techniques for the detection of printer roller pressure marks on paper. These pressure marks are due to the interaction between the paper feeder rollers and the paper substrate. ESDA was developed at the LCC, and its original application was in the detection of pen impressions several layers down in a notepad. It works by charging paper surfaces with a high voltage. If toner particles are applied to the charged paper surface, imperceptible pen marks due to writing on a piece of paper placed above this sheet may be revealed. This means that in a notepad, writing can be read from sheets many layers down from the top sheet that only has the actual writing. If the ESDA technique can detect weak imperceptible pressure marks from pens it might be able to detect pressure marks from printing rollers.

If successful, it will be more reliable and efficient than the methods described above for linking a document to the source that produced it. This is because the ESDA method is independent of potentially replaceable ink cartridges and text character shapes that are difficult to measure. Moreover, it offers the possibility of accurately determining the production source of a document even if a machine has to be selected from a group of machines that are of the same make and model.

If the pressure marks can be detected, to link a document to the printer that produced it would require comparing the width and spacing characteristics of the rollers of the machine in question. However, despite the simplicity of the methodology, detecting the pressure marks, if they actually do exist on the paper, is difficult for the following reasons. Firstly, the pressure exerted by the rollers is weak, generally much weaker than pen pressure. If the pressure marks are present, they will be difficult to detect, even with the most sophisticated digital image processing systems. Finally, heavy handling may destroy the marks.

The addition of computational techniques such as neural networks and Bayesian statistical analysis to the ESDA image analysis system can help detect the pressure marks. It has the potential to match weak, incomplete and noisy digital ESDA roller pressure image data to the rollers of the printing machine that produced it.

Jack Tchan
London College of Communication

used to support decisions regarding the allocation and priority of verification and validation activities. Having developed models for specific business sections, Rishabh described how they are now attempting to automate the process by building generic modules that can be simply slotted together.

The next presentation, by Vincente Grau of Oxford University, described some recent advances in real-time three-dimensional echocardiography. Unlike the conventional echocardiography devices used today that produce a two-dimensional `slice' through the heart, these new devices use a novel transducer to create real-time three-dimensional scans. In doing so, however, they create many new signal processing challenges, and Vincente demonstrated some algorithms to tackle them.

Finally, Ian Couzin of Oxford University rounded off the two days with a fascinating presentation of the real multi-agent systems found in nature; illustrated with a series of videos showing bird flocking, ants forming intricate trails as they foraged for food, and fish spontaneously forming emergent dynamic patterns. Many of these phenomena were the result of simple rules adopted by each individual and Ian showed how they could be reproduced in an agent-based simulation. Surprisingly, often changing a single parameter value was sufficient to flip the state of the entire system from one stable state to another. Ian's current work involves studying the formation and behaviour of locust swarms and he enlivened the talk with his experiences of living in a tent in Mauritania.

Overall, the Southampton meeting was a great success with two days of extremely interesting presentations and a memorable evening on the river Hamble.

Alex Rogers and Rajdeep Dash
University of Southampton

Since NCAF visited Southampton, one of the Electronics and Computer Science buildings suffered a major fire. Several of the attendees at the meeting will have worked in this building, and may have lost information as a result of the fire. NCAF hopes that their research will not be affected too much, and that any data lost can be recovered. **Nick Granville, Editor**

Natural computing and the National Measurement System

Most people in the UK know of the National Physical Laboratory (NPL) and maybe saw the recent celebration of the 50th anniversary of building the first atomic clock. Nowadays, the laboratory is privately run, but it is largely funded by the Department of Trade and Industry (DTI) as part of the National Measurement System (NMS). Much of their work is in support of manufacturing industry – aimed at improving industrial and scientific measurement in areas like engineering metrology, time and frequency measurement. However, it is perhaps less well known that the NMS also provides measurement support for the information and communications (ICT) sector and also has a Software Support for Metrology programme that has worked with neural networks.

The origin of the research programme in ICT was the DTI's Technology Review that identified the UK's key and emerging business technologies. The response of the NMS to this publication has been to set up a new research programme aimed at providing measurement support for emerging technology. This is known as the MET programme. There are five main MET research themes, aligned with the technology review – healthcare (mainly tissue engineering, implantable devices and diagnostics), energy (predominantly fuel cells), manufacture (nanotechnology, bioprocessing), materials (nanostructured and biocompatible) and ICT.

The ICT theme encompasses aspects of Pervasive Systems, Human Interface Networks, Complex Integrated Systems and Safety Critical Computing. At the moment there is one project under the ICT heading, 'Effective wireless sensor networks with context identification and communications.'

Wireless sensor networks are set to become a significant enabling technology in many areas of measurement, security and environmental monitoring. This project aims to develop an integrated approach to the metrology of sensor networks, including self-validation and calibration, network reliability and Quality of Service metrics. It will investigate the specific metrological and communication problems introduced by the wireless medium that affect the measurement confidence and efficacy. The work will include case studies and demonstrator networks and software to model and predict network and sensor performance.

As new themes and projects are identified they will be considered for MET funding, allowing rapid response to newly emerging technology. So, get your suggestions in now!

Moving to the more applied end of NMS work with software, a programme has been running for about eight years addressing Software Support for Metrology (SSfM). The first cycle of SSfM work looked at the then current practice in the use of software in measurement systems. There were some alarming discoveries. One version of the most widely used spreadsheet software calculated standard deviation in a way that accumulated rounding errors – not a problem for large deviations and small sets of numbers – but a significant cause for concern with large sets of numbers having a small standard deviation. NPL recommended robust algorithms for many common statistical operations and these are incorporated into the NAG library – and the spreadsheet problem has been fixed – for now.

More recently the programme has looked at the suitability of radial basis functions and support vector machines for modelling measurement related multivariate systems. The following is an extract from their conclusions:

"In view of the potential of RBFs to fulfil an important role in modelling and representing multivariate data, considerable research is being undertaken to improve computation times … and conditioning … as well as on the choice of centres … As RBF algorithms improve, they are likely to become a standard tool for multivariate modelling and data approximation in metrology. Much of the underlying technology of support vector machines – statistical information theory, kernel methods, etc., – are of potential value to metrology, particular in situations in which the system under study is imperfectly understood, for example, in biotechnology."

The current programme of software support for metrology includes the preparation of good practice guides in:

Visualisation and simulation
Discrete modelling and data fitting
Numerical software testing
Validation of software embedded in measurement systems
Data curation

For more information on SSfM see http://www.npl.co.uk/ssfm/index.html

For MET see http://www.npl.co.uk/met/index.html (subject to change so you may need to check with Google) or pass suggestions for future themes to Jessica.Rushworth@dti.gsi.gov.uk or met@npl.co.uk

Peter Cowley
Quarndon Cognition Limited
Peter is a member of the Measurement Advisory Committee and chairs its Quantum and Measurement for Emerging Technology Working Groups.

DIARY DATES 2006

13–16 February – AIA 2006: The IASTED International Conference on Artificial Intelligence and Applications, Innsbruck, Austria. http://iasted.org/conferences/2006/Innsbruck/aia.htm

5–8 April – ICCM 2006: 7th International Conference on Cognitive Modelling, Trieste, Italy. http://iccm2006.units.it/

10–12 April – EVOBIO 2006: 4th European Workshop on Evolutionary Computation and Machine Learning in Bioinformatics, Budapest, Hungary. http://www.cs.vu.nl/~evobio/evobio06.html

May 2006 – NCAF meeting, venue, dates and theme yet to be announced.
For information, email enquiries@ncaf.org.uk or telephone +44 (0)1332 246989.
29–31 May – ISNN 2006: 3rd International Symposium on Neural Networks, Chengdu, China. http://www2.acae.cuhk.edu.hk/~isnn2006/

MEMBERS' NEWS AND VIEWS

Deadline for contributions for the next edition – 1 February 2006. Please send to Managing Editor – Nick Granville, e-mail: Nick.Granville@smith-nephew.com