ncaf Natural Computing Applications Forum

Networks

Industrial Applications in the Industrial Midlands

University of Birmingham, 14–15 June 2006

Real world applications can be a great inspiration for researchers, presenting new and complex challenges.

I am pleased to invite you to attend our next NCAF meeting, to be held at the University of Birmingham. Our theme is the 'Industrial Application of Natural Computation'. Nature inspired techniques are becoming more and more common in areas as diverse as product design, manufacturing and prediction, and we plan to have some very interesting speakers from both academia and industry.

The meeting is hosted by Cercia (the Centre of Excellence for Research in Computational Intelligence and Applications), a group founded by Professor Xin Yao at the University of Birmingham. The overlap of industrial and commercial applications with academia is a key focus for Cercia, who have been involved in the organisation of the workshop 'Applications of Evolutionary Computation in Business' at CEC 2005, a recent workshop on 'Computational Intelligence in Cheminformatics' and the forthcoming CEC 2006 workshop on 'EC at Work - Generating Value with Evolutionary Computation'. Real world applications can be a great inspiration for researchers, presenting new and complex challenges. For those working in the commercial world, these workshops provide a great opportunity to find out about new research and meet potential collaborators.

Speakers confirmed so far are Ian Stott of Unilever, who will speak on the use of genetic algorithms to design biologically active molecules, and how the company have obtained patents using neural networks. Dan Lin from the University of Birmingham will speak on the evolution of routes for road salting trucks. We also hope that Professor Dave Cliff of the University of Southampton will speak.

There are a few speaker places still available, so if you would like to present at the meeting, either on industrial applications or more generally on natural computation, please contact me at the address below.

Cercia works with businesses to apply computational intelligence (CI) techniques to real world problems. It is based in the School of Computer Science, and draws on the work of more than 25 full time researchers in CI. Work in the group includes data mining, analysis of biological signals such as EEG, route optimisation and evolutionary design. The group also makes use of 3D visualisation techniques for data analysis. Less applied work includes research into the mathematic properties of evolutionary algorithms and the design of neural network ensembles.

Cercia works with both regional and international industrial partners including BT Exact, Honda Research Institute, HP Labs, Marconi, Qinetiq, Rolls Royce, SPSS (UK), STMicroelectronics and Thales Research and Technology. Cercia also has funding from the regional development agency to work with small and medium sized enterprises. This includes direct consultancy and training courses in areas such as data mining and evolutionary computation.

Birmingham is a campus university and I hope that the weather in June will be fine, allowing us to enjoy the University Botanical Gardens which are adjacent to the conference centre. The gardens house the National Rose Collection, which should be in full bloom, providing 'inspiration from nature' for NCAF delegates. Bowling proved to be a hit in at the last NCAF meeting in Swansea, so we have decided on dinner followed by a re-match as the social event.

In conclusion, NCAF meetings are a great opportunity to find out what is happening at the cutting edge of applied research. I guarantee that you will go away buzzing with new ideas!

Andy Pryke
Cercia, University of Birmingham
Email: A.N.Pryke@cercia.ac.uk

INSIDE

BACK PAGE

Data mining sparks im

Temblet acknowledged that data mining has considerable potential for the health service … as a diagnostic tool and for monitoring public health.

NCAF came to Wales again for the winter meeting and the delegates enjoyed the two days of stimulating talks on data mining. The meeting was co-sponsored by the Civil and Computational Engineering Centre and IT Wales.

Making sense from previously known and probably incomplete data is not a trivial problem. The delegates experienced the need of better data mining models immediately on the arrival in Swansea. They had to locate the Premier Travel Inn and its neighbouring NCP Car Park on a dark, wet night and the only data they had were the directions printed on the Premier Travel Inn's website.

On the first day, the meeting kicked off with a presentation from Steve Woods of Informavores describing the process of decision automation and knowledge encapsulation, an essential part of their products and data mining. Using an example of complex and multi-faceted international business, Woods described the difficulties of assigning tax codes where complex rules and changing regulations affect everything. Firestorm software from Informavores allows the user to design the process as a flow chart, or 'spark', with questions and answers, without the need for programming. This thought process is mapped using diagrams, and the working piece of software is 'drawn' by the user. By applying procedures to data, Woods demonstrated clearly that costly mistakes can be avoided, and time can be saved using an extremely powerful, yet simple way to analyse vast amounts of data.

After this presentation, our host, Rajesh Ransing stepped in to deliver a case study presentation on behalf of Ontrack Systems, which described an opportunity to use data mining to save millions of pounds for the NHS. Each GP appointment costs the NHS £18, and in one month of 2005, research showed that some 10 million appointments were missed. The presentation maintained that effective communication is key to this issue, utilising voice and SMS along with other methods to follow up and remind patients. Using the patient data, the NHS could communicate with patients via letter and SMS, giving the patient the opportunity to respond and interact directly. Such a project would require the implementation of a two-way SMS gateway.

3D Views

Simon Hickinbotham of the University of York described high performance pattern recognition technology which can search databases for 3D views of complex molecules. Hickinbotham explained that the AURAmol technology allows users to search for molecules that match a particular description, or define the similarities between them to enable the system to use this knowledge of similarity to locate others automatically. A demonstration of the software over the Internet is planned for 2006.

Professor John Tucker of the University of Swansea discussed data organisation and requirements for SMEs including the need to hold a wide variety of data for short term use, long term audit and offsite security. Tucker defined data mining as being part of data centric computing, a theory for computation where the data are viewed as the primary concept upon which a variety of methods including programming languages, databases and visualisation can all be built. Tucker described how programming systems could be represented theoretically, and maintained that algebraic modelling and algorithmic simulation were extremely significant in this.

Professor Peter Flach of the University of Bristol discussed the use of Receiver Operating Characteristic analysis to improve data mining models. Using the example of radar, Flach demonstrated that once an object is detected, the radar must work out what the object is based upon the characteristics of movement and size. ROC analysis is a method of analysing data by predictive supposition, where the solution used learns from previous correct and incorrect responses to analyse data in a more educated way.

The first day of the conference came to a close with a presentation on the need for data mining in the NHS. Paul Temblet, a doctor working in Llanelli, South Wales, painted a grim picture for delegates when he stated that 1,200 people die each year in hospitals in England and Wales due to medical errors brought about by incomplete or inaccurate data. These errors cost the NHS £500 million per year. The problems are caused by an inefficient paper based record system, increasingly complex diagnoses and more complicated treatments. Dr Temblet gave the example of a cancer patient requiring hospital admission whilst away from home, but without access to his medical notes for 48 hours, the patient's condition deteriorated and he finally died. The current paper based system can have a serious impact on patient care, but as Temblet stressed, the present use of technology does not give much scope for improvement with hospitals and GPs practices using discrete databases with no way of transferring data between them. Temblet acknowledged that data mining has considerable potential for the health service including highlighting trends or disease outbreaks, as a diagnostic tool and for monitoring public health. He is working with the University of Swansea to develop self-learning software to assist medical staff undertake and improve diagnosis and response. In conclusion, Temblet said that the IT systems of the NHS must be "dragged into the 21st century".

The day was also packed with entertainment. A coach trip to the newly established National Waterfront Museum was arranged. Wales was one of the earliest and most heavily industrialized nations on earth. The delegates saw how Wales and her people influenced the rest of the world throughout the Industrial Revolution. The social evening combined ten-pin bowling with an Indian meal. NCAF Chairman, Graham Hesketh indulged himself in data mining, coming up with lot of statistics on the next day. He entertained the audience with best and worst performances of groups/individuals with different criteria and also gave feedback based on ROC analysis.

Collaborative research networks

The second day of the meeting began with insight into the ways in which data mining can stimulate business growth. Colin Frayn of CERCIA at the University of Birmingham demonstrated force

nagination at Swansea

based visualisation of data which provided accurate and instant analysis of the various properties based upon the parameters set by the user. The CERCIA technology is helping businesses in the West Midlands develop an appreciation for the importance of their data and understand how data mining can encourage business development. It is also used to map collaborative research networks in local schools to draw together those with similar interests. Frayn went on to explain the ways in which force based visualisation could be used to map both computer and human networks, to predict or show failures. He also discussed the use of the techniques in e-commerce websites to enable webmasters to analyse customer routes through the site and extrapolate patterns. Future developments for the technology will include 'active stereo visualisation', a 3D visualisation system which will make the data mining process both more intuitive and interactive.

Former police officer Rick Adderley of AE Systems described how he was using data mining techniques to "drag the police force into the 21st century". Police data mining does not deal with vast amounts of data but relates to data from several police forces, some dealing with over 1,000 reported crimes per day and more than 750 intelligence reports. Much of the police data is paper based, with reports taken by police officers, in free text, and entered into databases by civilian data entry clerks. Once in the database, the data can be used by analysts, but there are often inconsistencies in the analytical process with varying perceptions and interpretation of the data. Adderley went on to discuss the potential of data mining in the investigation of serious crimes to cut the costs associated with time consuming physical data searches, and also the prioritisation of data exploration of crime scenes by forensic examiners. In conclusion, Adderley summed up the other potential uses of data mining in operational policing, including professional standards and HR.

Meghana Ransing presented X1Recall, a diagnostic tool developed at the University of Swansea to help manufacturers produce defect free components. The patented software solution uses a knowledge base of past analyses to enable it to correct and rectify errors. Meghana explained that the solution accepts feedback and learns from its own mistakes to enable the production of perfect products.

David Lowe of the University of Aston discussed steganography, a method of hiding data within data. This is typically associated with downloaded files such as MP3s and images, and plans are being put in place to allow the file owner to link the downloaded version of the file with the purchaser, so that if it is illegally passed on, the original person can be identified. The individual purchasing the MP3 will have no idea that the file has any other data encoded within it, as the sound quality remains the same. Lowe went on to discuss the concept of digital watermarking in medicine, where it is easy to hide data, such as patient records, within large biopatterns including MRI scans. While the scan would be modified slightly to incorporate the hidden data, it would not impact on the medical staff analysing the scan.

Just before lunch, Puzzle Corner was solved in the usual dramatized way. The AGM followed lunch, after which there was the final presentation. Nazri Nawi of the University of Swansea discussed training algorithms in neural networks and showed that convergence can be improved by choosing a better search direction based on the adaptive gain variation. Nawi demonstrated that the method under development is faster than the standard method of back propagation, while providing results with significantly increased accuracy.

Overall the Swansea meeting was a great success. It was packed with excellent, high quality presentations covering a range of data mining topics from theory to applications, a memorable visit to the museum and an evening full of fun and activity. We hope that NCAF will not hesitate to come back to Swansea again in the future.

Sali Earls, IT Wales
Rajesh Ransing, University of Swansea

NCAF AWARD

for the best postgraduate student presentation

The second day of the Autumn meeting will be given over to student presentations, and will be a showcase for student research.

This award is a new opportunity for students who are expecting to submit a Masters' thesis (MPhil, MRes or MSc) in 2006, or have already submitted their thesis in 2005 to any British university. The thesis should be in the field of natural computing or of interest to NCAF members.

A 600-word summary and the slides that the student intends to use for the presentation must be submitted by 1 August 2006. The NCAF committee will invite a subset of the students to present at the NCAF meeting, should there be too many submissions to accommodate at the meeting. There will be a prize for the best presentation, as chosen by the non-student attendees at the meeting.

Further information on the competition and submission details will be available on the NCAF website from June 2006.

PUZZLE CORNER

Number 32

Waking up in Hotel Morrison, Lisa found herself in the infamous Manzarek Suite. The bedroom had 7 doors (conveniently labelled A-G in big easy-to-read letters). Doors A-F all joined the Suite to an adjacent antechamber, but the antechamber had no other exits. That is, she could get into the antechamber from her bedroom, but the only way out of that room was back into the Suite through one of the other doors. Door G, however, did lead to freedom (and breakfast). Unfortunately it was locked, with 6 electronically operated bolts, labelled AF. Currently all the bolts were shut, but some helpful instructions were written beside the door.

'3 of the doors A-F can only be opened from the bedroom; the other 3 can only be opened from the antechamber. Opening one of those doors automatically locks all the others. Closing the door, which can only be done after passing on through (to the other side) will unlock some of the other doors AND *change the state of the corresponding bolt on door G. Closing door A unlocks doors B&E (and toggles the state of bolt A); B unlocks C&E; C unlocks D&F; D unlocks A&F; E unlocks B&D; F unlocks C&D.'*

Lisa could see that doors A and F rather helpfully had 'Opens from other side' signs on them, and door B had an 'Exit this way' sign. The other doors were unhelpfully unsigned. Nevertheless, Lisa quickly figured out how to negotiate the doors A-F in the shortest time possible and was down to breakfast before the chef could say "Light my fire, Gloria, it's time to open the sole kitchen!" Strange days.

What was the sequence of doors that Lisa went through?

The answers will be given at the next NCAF meeting (14–15 June 2006, University of Birmingham).

Fenella the Rottweiler

ncaf Natural Computing Applications Forum

COMMITTEE NOTES

MEMBERS' NEWS AND VIEWS

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

Edited and Produced by:
Forum Communications
Westgate House, Old Ivy Lane
West End, Southampton
Hampshire SO30 3RX
Tel/Fax: 023 8047 6888
e-mail: info@forum-pr.co.uk

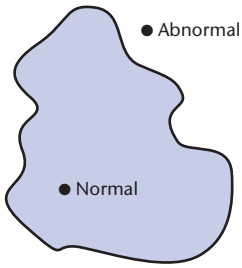
NEXT EDITION
Review of Birmingham meeting
Preview of Bristol meeting

Intelligent monitoring of high-integrity systems

Oxford BioSignals employs techniques of novelty detection, data fusion, and visualisation for condition monitoring of complex, high-integrity systems. Successful implementations to date include fields as diverse as the monitoring of patients' vital-signs in hospital Intensive Care Units to the monitoring of jet engines in the latest generation of Eurofighter, Airbus A380, and Boeing 787 aircraft.

In each case, advance warning of unusual behaviour enables pre-emptive actions to be taken to avoid system distress, allowing use of a pre-emptive system maintenance strategy.

The nature of many high-integrity systems is that there are few examples of abnormal operation and failure, due to the rarity of such events in comparison to the large amount of available data representing normal system operation.



The novelty detection paradigm suits such applications, in which normal system behaviour is modelled. Significant deviations from a model of normality are identified as potential hazards, or precursors of hazards. This approach allows meaningful prognosis and diagnosis of abnormal system conditions for which representative fault data are unavailable.

In producing a monitoring solution for a target system, Oxford BioSignals use a variety of soft computing methods for novelty detection. Initial data studies of a new system are typically undertaken to demonstrate 'proof of principle' in applying condition monitoring techniques to the target domain. Through partnership with domain experts, the training of algorithms and construction of models of normality is guided by exploiting *a priori* knowledge of system behaviour.

It is usual for high-integrity systems to be monitored by a variety of sensor types. A monitoring suite for a human body could include blood pressure, body temperature and heart rate; a similar system within a jetfighter engine could include vibration measurements, oil temperature and air pressure.

Conventional monitoring systems are often limited to examining each type of signal independently. However, better monitoring performance may be achieved by a holistic consideration of different sensor types. Such data fusion techniques can allow better predictive and diagnostic power than conventional methods. Fusing information from various sensors in order to perform novelty detection is a substantial benefit of the soft computing approach. A particular advantage of such techniques is the capacity for monitoring systems to generalise from training data to previously unseen examples. In creating a system to monitor the health of a ship's engine, for example, a training set may consist of many instances of engine operation from a variety of different engines of the same type. From those different engines, a generalised model of normality for the whole class of engines is learned. The resulting model may be used for condition monitoring of any engine of that class – though engine-specific methods can be employed to overcome inter-system variation if sufficient training data are available.

When modelling high-dimensional systems, in which a large number of different sensors are employed, producing visualisations of that system in a manner that is meaningful to a user is of paramount importance. Only limited information can be determined by displaying sensors in isolation, or in simple X-Y scatter plots. Visualisation allows the evolution of the system's behaviour to be examined over time. This is frequently used both during the training stage of novelty detection, and in actual service.

In visualisation, a projection model is trained upon the data to be visualised. The input set of sensors is known as the input space, and typically may contain many dimensions, perhaps over 20. These are then reduced by the visualisation network to a lower number, typically 2 (so that they may be plotted on 2-D axes for human inspection).

The training process causes the projection model to preserve relationships between the data in their original high-dimensional space, such that two patterns which are close together originally are still close together when projected onto 2-D axes. The result is a 2-D plot of the system's operation over time, with each sample of system state represented as a single point.

This method is extremely powerful in showing the evolution of a complex system over time. A system may stay within one particular area while normal, and then gradually move away from that region as the operation becomes more and more abnormal. Application of corrective maintenance actions frequently results in the system returning to the region of the plot corresponding to normal operation.

In being able to examine sets of high-dimensional data, the knowledge of domain experts can be used to determine areas of the visualisation corresponding to particular modes of behaviour. One region of a jet engine's visualisation might correspond to the failure of a bearing, whilst another area could indicate failure of an oil subsystem.

The techniques introduced within this article form part of the array of methods used by Oxford BioSignals in creating monitoring solutions for complex high-integrity systems. Although each application is constructed separately to be most appropriate to the domain of application, the same general data-driven methods may be applied to almost any target asset.

David Clifton
Oxford BioSignals, and Signal Processing Research Group, University of Oxford

DIARY DATES 2006

Mid-September – NCAF meeting at Bristol. Dates to be announced, and including the first NCAF Award for the best postgraduate student presentation.
For information, email enquiries@ncaf.org.uk or telephone +44 (0)1332 246989.

11–13 December – AI-2006: 26th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence.
Cambridge, England
http://www.bcs-sgai.org/ai2006