A new field in an old city

Bioinformatics at York, University of York, 10-11 April 2001

The April meeting of NCAF will take place at the University of York with a first day theme of bioinformatics. The York area has one of the largest concentrations of biotechnology companies in Europe and the University is at the forefront of the bioinformatics revolution, playing a part in, amongst many other things, sequencing the first plant genome. For those of you who are a little hazy about what exactly bioinformatics is all about, the morning will start with a tutorial session to bring you all up to speed. This tutorial will be given by Sandie Baldauf (Department of Biology, University of York).

This will be followed by Rod Hubbard of the York Structural Biology Laboratory (Department of Chemistry, University of York) who will discuss recent advances made in the analysis of protein structure and function. The Structural Biology Laboratory has a significant interest in analysing structures solved in York and elsewhere, both to rationalise the structural features and to predict and design new molecules.

The morning concludes with Mahesan Niranjan (Department of Computer Science, University of Sheffield) who will be talking about machine learning tools for the analysis of microarray gene expression data.

In the second half of the day, Stephen Muggleton (Department of Computer Science, University of York) will start off by talking about knowledge mining in the chemical and biological domains. He will give us some valuable insights into the ways in which the raw chemical and biological data can be turned into useful information. The talk will review the results of the last few years' academic pilot studies involving the application of Inductive Logic Programming to the prediction of protein secondary structure, mutagenicity, and structure activity. Following this, Mike Sternberg (Biomolecular Modelling Laboratory, ICRF) will discuss some of the ways in which the computational tools developed to deal with the vast bioinformatics data sets can be applied to analytical genomics. The day concludes with a presentation from Elizabeth Huddleston (Smith + Nephew) on the ways in which bioinformatics has had an impact in the healthcare markets that Smith + Nephew are involved in.

The evening's social event will be a welcome repeat of the mediaeval banquet that was so well received at the last meeting in York. This will be held in the Merchant Taylor's Hall just inside the ancient city walls, and entertainment will be provided to complete the mediaeval atmosphere.

The second day covers a broader range of topics than the first, but still retains a flavour of the first day's theme. The first speaker is Jim Austin (Department of

Computer Science, University of York) who will be talking about high performance pattern recognition. The technology is applicable to a wide range of domains, and applications include the high speed searching of molecular databases. Arun Holden (Computational Biology Group, University of Leeds) will follow, talking about the challenges involved in the computational reconstruction of the heart. The final talk of the morning will be from Joab Winkler (Department of Computer Science, University of Sheffield) who will give a tutorial on the use of wavelets.

The meeting will be rounded off by Gerry Mogg (Department of Trade and Industry). Gerry will be describing the DTI's Management of Information programme. The programme is aimed at companies, particularly smaller firms, operating in the ICT sector who wish to collaborate with an academic partner to develop technologies and systems that manage and process information, especially information held in electronic format, for the control of fraud and the improvement of security and privacy.

Simon O'Keefe
University of York

Aston reviewed
Review of Aston meeting, 24 - 25 January 2001

NCAF's January meeting started with Keith Copsey, DERA, who gave us a gentle introduction to particle filters from their beginnings in the 1970's to present day state of the art use. Now general availability of sufficient computational power allows them to be of routine practical use. Keith rather bravely included real time online demonstrations to show their advantages over extended kalman filters with application to trajectory tracking with multiple targets and obscured regions. His bravery was rewarded when the demos actually worked!

The time series theme continued with a call to arms from David Lowe of Aston University. He first of all set the scene by tracing the history of time series analysis from the 1900's to the present day. There was a fundamental split into the two camps of a dynamical systems approach and an approach based upon stochastic processes. David then took us on a tour through each of the two camps using many examples such as sunspot activity, monthly tourist figures, electricity load demand and financial series. Insight was provided into the two different philosophies and neatly concluded by showing how with a proper perspective both approaches had a similar underlying view and it might be possible to reunite them in the future.

Sebastian Larsson (Econostat) provided an overview of the principal activities undertaken by Econostat, who provide consultancy related to predicting medium to long term economic trends and also short term portfolio services related to single month moves in the bond and equity markets. In general the main problems with financial time series were cited as sparse noisy data, large number of variables, inadequate a priori information, and unknown dynamical relationship between variables. This set the scene for Neep Hazarika (Econostat) who reviewed the various techniques used to construct Econostat's portfolios and models. Neural

networks (except Bayesian) were largely deemed unsuitable because of scarcity of data. In practice it was considered better to use univariate or multivariate models. Econostat are also looking at investigating other domains for insight into new approaches e.g. combustion equations for 'bursty' time series, dynamic time warping from speech recognition and singularity modelling.

Genetic Algorithms

There followed a brief departure from the time series theme of the first day to bring us an excellent talk from Xin Yao (Birmingham University) who gave an overview of Genetic Algorithms with examples derived from the two player prisoners' dilemma (and n-player) to try to examine how optimal strategies are arrived at for wider application elsewhere. Xin demonstrated the impact of history length, impact of group size and how reputation encourages co-operation more than any other attribute. He provided insight into co-evolution and co-operative behaviour and why this is a nice approach to learning since solutions can be evolved from random strategies without an explicit fitness function and therefore without expert knowledge or a teacher.

Tom Harris's (Cardionetics Ltd) talk began with an overview of Cardionetics and its key product, the C.Net2000 ambulatory ECG monitor and event recorder. Resting ECG only indicates problems with very acute cases and the main alternative is a 24-hour hospital based study which is unnecessarily expensive and typically has a long waiting list which can be up to 40 weeks. The problem is also exacerbated by the imprecise and vague symptoms described by the patients. Often serious cases, which would ideally need swift referral for treatment, are embedded in a queue of people who have nothing wrong with them at all (between 40%-60%). During their wait stress may cause a deterioration in their condition, which then requires emergency treatment.

The key difference between the C.Net2000 and its competitor's products is that all the analysis is performed in the instrument rather than offline at a central location. Analysis is performed by a committee of Kohonen SOMs which are trained offline, but in real time classify each heartbeat as it occurs.

The talk continued with Ian Nabney who is Director of the recently formed University Cardionetics Institute of Bioinformatics. Ian outlined the role of the Bioinformatics Institute which effectively forms the long term research arm for Cardionetics. It is apparent from an initial survey that methods and algorithms employed in current EEG work is lagging other allied fields by several years.

The reception (sponsored by Cardionetics) allowed NCAF members a free check up with a C.Net. Several delegates were spotted running up and down the corridors to see just how high their heart rate would go! The day concluded with a well attended evening of ten-pin bowling followed by dinner at a tapas bar. Many people claimed that their aim at bowling was improved by the supply of champagne, though no significance tests were carried out to verify this.

Spin models

It was not necessary to involve sophisticated modelling techniques to deduce that day two dawned a little too early for some enthusiastic participants of the previous night's social event. Nevertheless the day got off to a bright start with an entertaining talk from David Saad (Aston University). David reviewed his ongoing work applying statistical physics to the area of cryptography and error correcting codes. After a brief overview of the need for such techniques in applications such as satellite communications and magnetic media storage he went on to show how the application of spin models could achieve worthwhile improvement in cryptography over the commonly used RSA algorithm yielding advantage with decryption times and correction of mild corruption but at the expense of needing a longer public key.

Chris James (Aston University) gave an overview of some of the work being undertaken within Aston's Neurophysiology Applications Group relating to epilepsy and brain source localisation. This entails having to process a very large amount of data from a magnetoencephalographic (MEG) scanner. This type of scanner allows a view of the activity of the millisecond by millisecond changes of activity within the brain's cortex but is characterised by a large number of data channels (64-155) of very low amplitude signals containing a large number of artefacts. He then went on to review the techniques being employed to automate the current manually intensive detection of epileptiform discharge (spike) detection which is of importance in epilepsy and also localisation of activity within a region of the brain. Epilepsy is often treated by the removal of an abnormal region of the brain so it is crucial to detect the earliest signs of a seizure to help in the accurate determination of its locality before the brain becomes swamped in a storm of activity.

Peter Tino (Aston University) gave a presentation of the work he is undertaking with regard to visualisation which is a useful aid in pharmaceutical drug discovery. He has built on prior work performed at Aston so that a hierarchical tree of visualisation plots can be built up in an interactive manner to aid meaningful interpretation of the data.

Hujun Yin (UMIST) reviewed the early beginnings of independent component analysis and its application to blind source separation, demixing and source separation without a priori knowledge. In a nice demonstration he showed how multiple speakers in different languages could be well separated with ICA.

David Barber (Aston University) wrapped up the meeting in fine style by allowing us to observe the differences in data that can be produced with changes to the topology of graphical models. It was a graphic (or rather aural) demonstration that was impossible to ignore since it would appear that graphical models have a penchant for rather loud drum music. This brought another excellent NCAF meeting to a close.

Michael White
Aston University

Forecasting property

NCAF stalwarts Chris Satchwell and Howard James have recently set up a new enterprise. Howard describes the processes and challenges involved.

When we decided to set up a company to forecast property prices (Technical Forecasts Limited), we had a feeling we had chosen a difficult task. As time went on, our initial feelings became fully justified. The neural side of it was not exactly routine, but it was solvable; the real task was the data. In order to forecast ahead, we needed good data going back in time.

On the face of it, the data looked reasonable. Past property data series do not all have the volatility of stock market series; property movements are more cumbersome than stock market series because price changes take time to work through the system so there was a chance that our work would not be bedevilled by wild fluctuations.

The real challenge was the shortness and sparseness of property data series. They only have figures once a quarter, or in some cases, once a year. To make it worse, some of the important series are very short; for example, Land Registry house prices only go back to 1995. We discovered early on in our investigations that auto-regression time series forecasts were too inaccurate to be useful. Neural networks gave some advantages, but still the reliability was poor. It was clear to us that the short and sparse property series that were available did not contain sufficient information to allow the series to be projected into the future.

It was therefore natural to turn to multiple series. It seemed a good idea to add information to the property series from other series that were in some ways related to the property series being forecast. The challenge was to find the right series.

Some earlier work on a series of commercial property prices showed that better forecasts could be obtained by using a parallel series of gilts (British Government Securities). Clearly, the gilts added information that was not present in the past series of values. At last, we were seeing a breakthrough.

Since that time, the technique of using parallel and multiple parallel time series data has been developed so that now the techniques are used routinely to forecast residential and commercial prices (capital values, rents and yields).

For example, it is possible to forecast features in a property time series. A prediction using a parallel series is shown in the forecast above. The real series (diamond points) hesitates before continuing its upward trend. The ANN forecast (circles), from May 1998, indicates some instability but then picks up the trend of the series. This forecast was impossible without the parallel series, even though the parallel series did not contain the same feature that was in the main series being forecast.

The series being forecast was an Insignia Richard Ellis series of capital retail values, and the parallel series in this case was another property series. A paper describing this work, by James and Connellan, is available at www.rics.org.uk from the proceedings of the RICS Cutting Edge Conference of September 2000.

Howard James

PUZZLE CORNER
Number 16

In a fit of excess energy, Lisa signed up with the Territorials just in time for their MMM outing (Midnight Manoeuvres in the Mud). This session was billed as an Exercise in Problem Solving using only Regional Communications.

As Lisa discovered, a large field had been marked out as a 4x4 grid of sectors designated A-D (running West to East) and 1-4 (running North to South). Her platoon consisted of four squads: Hawks (starting in B2), Kestrels (in B3), Buzzards (in C2) and Falcons (in C3).

During each 5 minute time slot only one squad is in charge. That squad is responsible for ordering only one orthogonally adjacent squad to move. The move must be in one of the four cardinal directions (N,S,E, or W) and the moving unit must continue in that direction until they can go no further (they may not leave the field, and they may not enter a sector already occupied by another unit). Units move at a rate of one sector per minute. No squad may move unless it is ordered to do so by the squad currently in charge.

The exercise begins at 12 noon. Every 5 minutes command updates according to a schedule, and the next movement order can then be issued. The group objective is to get the four squads into the corner sectors of the field (A1, A4, D1 and D4), at which point they can simultaneously deactivate the electric perimeter fence. A secondary objective is to do so before the anti-personnel mines in the field are remotely activated at 12:55 precisely.

The platoon has little faith in their commander. They have asked Lisa to draft a plan (i.e. command schedule and orders) to get them all out alive.

What was Lisa's suggested command rotation and associated movement orders? The answers will be given at the next NCAF meeting (10-11 April 2001, University of York).

Fenella the Rottweiler

Cybula high performance pattern recognition

Cybula is sponsoring this April's NCAF meeting at York. The company has grown out of Jim Austin's group at the University of York, known for some time for its leading research on the AURA binary neural network. The group has been developing systems for many commercial organisations, such as British Aerospace, The Post Office and Glaxo-Wellcome. Now, with a proven technology, the systems

will be fully supported and marketed. Cybula delivers high performance pattern recognition systems that deal with very large amounts of data accurately and quickly. Its first products are based on the AURA neural network as a search engine and are offering a text retrieval system for address databases. This allows a user to search customer address files using only fragments of the address. The second system is a trade mark image database, which allows companies to find out if a trade mark image has already been registered by other companies. Both systems operate on many millions of items and are supported by a generic high performance AURA search engine running on computing clusters. More details can be found at

http://www.cybula.co.uk and by contacting Cybula on   01377 236 382.

Simon O'Keefe
University of York

Ila takes a break

"After being the Managing Editor of Networks since 1994 I am finally stepping down. I would like to thank all the contributors who have sent in articles over the years. Thank you for remaining polite to me in spite of my constant nagging and cajoling. The efficient turnover of the newsletter is mainly due to the combined efforts of Sally and Chris Hawthorne at Forum for which I am extremely appreciative. I finish by wishing Rick Adderley the best of luck in taking over from me and hope that all our members will co-operate in supplying future articles"

Ila Patel

New Editor

"Greetings to all readers. Having been an NCAF committee member for the past 12 months, I now have great pleasure in taking over the editorial role for our Networks newsletter.

I am a serving Police Officer with the West Midlands Police currently using various types of neural networks within the envelope of datamining with a view to providing analysis on crime trends and offender profiles. I would like Networks to continue in a similar vein at present and solicit your views as to its future direction.

I would like to take this opportunity to thank Ila Patel for her hard work, persistence and patience, as the current editor, and hope that I can meet the high standards that have already been established."

Rick Adderley
West Midlands Police

DIARY DATES 2001/2

10-11 April NCAF, University of York
Contact: Mark Cheeseman e-mail: mark.cheeseman@rolls-royce.com Tel: +44 0 (1332) 246989

13-15 June Workshop on Self-Organising Maps 01, Lincoln, UK --
http://wsome01.ee.umist.ac.uk

13-4 July NCAF, University of Oxford (Wadham College). Contact: Mark Cheeseman e-mail: mark.cheeseman@rolls-royce.com Tel: +44 0 (1332) 246989

20-21 September NCAF, Royal Holloway College. Contact: Mark Cheeseman e-mail: mark.cheeseman@rolls-royce.com Tel: +44 0 (1332) 246989

4-6 September 14th International Congress on Condition Monitoring and Diagnostic Engineering Management (COMADEM) 2001
http://www.eng.man.ac.uk/mech/comadem.htm

10-12 December ES2001: Cambridge, England, http://www.bcs-sges.org/es2001/
Closing date for submission of papers is 25 May 2001. Contact Richard Ellis e-mail: richard.ellis@stratum-management.co.uk

January 2002 NCAF, University of Exeter. Contact: Mark Cheeseman e-mail: mark.cheeseman@rolls-royce.com Tel: +44 0 (1332) 246989

COMMITTEE NOTES

Chairman
Graham Hesketh
Rolls-Royce plc

Secretary
Mark Cheeseman
Rolls-Royce plc

Treasurer
Dr Andrew Starr
University of Manchester

Editor of Neural Computing & Applications Journal
Professor John MacIntyre
University of Sunderland

Managing Editor of Networks
Inspector Rick Adderley

West Midlands Police

Publicity Officer
Richard Ellis
Stratum Management Ltd

Mandy Bradley
Royal & Sun Alliance Insurance

Dr Simon Cumming
British Airways plc

Dr Richard Everson
Universty of Exeter

Dr Neil Lightowler
AXEON Ltd

Dr Ian Nabney
Aston University

Dr Kathryn Burn-Thornton
University of Durham

Please contact NCAF through Mark Cheeseman
Secretary - NCAF
PO Box 111, Manchester
M13 9QF U.K.
Tel: +44 0 (1332) 246989
Fax: +44 0 (1332) 247129
e-mail: mark.cheeseman@rolls-royce.com
http://www.ncaf.co.uk

Retiring Managing Editor:
Dr Ila Patel
17 Millpond Court
Addlestone
Surrey KT15 2JY
e-mail: ilapatel_2000@yahoo.com

Edited and Produced by:
Chris Hawthorne
Forum Communications
Eastgate House
Eastgate Street
Winchester SO23 8DZ

Tel: 01962 877833
Fax: 01962 877988
e-mail: chrishawthorne@forum-pr.co.uk

MEMBERS' NEWS AND VIEWS

Deadline for contributions for the next edition - 11 May 2001 Please send to: Managing Editor, Inspector Rick Adderley, e-mail: Insp1908@aol.com

NEXT EDITION

Preview of the Oxford meeting
Review of the York meeting