ncaf Natural Computing Applications Forum

Networks

Signal processing and medical applications amidst the dreaming spires

University of Oxford, 18-19 June 2008

The theme of this meeting is signal processing and its application to biomedical engineering

NCAF returns to Oxford after a seven-year interim for the Summer Meeting, which will be hosted by the Department of Engineering Science. Reflecting the Forum's twin strengths of providing an opportunity for discussing new theoretical methods and their application to real-world engineering challenges, the programme consists of invited speakers from industry and academia. As ever, the aim is to provide all this within the familiar friendly NCAF setting, one of the main meetings of the UK community in Information Engineering and related disciplines.

Following the success of the last Aston meeting, where we focused on Machine Learning techniques, the theme of this meeting is signal processing and its application to biomedical engineering. The latter theme represents one of the fastest-growing interests in modern engineering, and is a particular speciality at Oxford with the recent opening of the University's Institute of Biomedical Engineering (IBME). The Director of the Institute and long-time supporter of NCAF, Professor Lionel Tarassenko, will provide the first keynote presentation on some of the novel research being undertaken at the IBME.

Describing new directions in machine learning, with particular relevance to signal processing and biomedical science, Professor Christopher Bishop of Microsoft Research will present examples including genomic analysis.

Coupling the discussion of both signal processing and examples of its application, Dr. Christopher James will be describing recent advances in biomedical signal processing made at the University of Southampton, while Dr. Hujun Yin will introduce research from the University of Manchester, focusing on his group's strengths in machine learning and biologically-inspired processing. The use of information engineering techniques to solve a host of other problems will be presented by Dr. Stephen Payne and Dr. Nick Hughes, both of the IBME.

Firmly convinced that the best engineering is informed by real-world needs, one of NCAF's priorities is the presentation of successful commercial applications. As an example of this, Dr. Iain Strachan will describe the use of probabilistic time-domain analysis for tackling commercial problems within the context of his own research at university spin-out company, Oxford BioSignals.

David Clifton, University of Oxford

Radcliffe Camera, Oxford

Dr. Jiwon Yoon of Oxford's Pattern Analysis Research Group will tackle the area of Bayesian inference for providing a Brain-Computer interface.

Apart from the discussion of cutting-edge research and its industrial application, one of NCAF's primary reasons for existing is to provide a place to meet the UK's 'Natural Computing' research community in an informal and friendly atmosphere. NCAF prides itself on being the place where lasting professional and personal friendships are formed. After attending just a few NCAF meetings, one truly begins to see how close-knit this community is. So, we take 'networking' very seriously. So seriously, in fact, that this meeting's social event will begin with a barbecue in the picturesque setting of an Oxford college, followed by the traditional punting-and-Pimms up the river, accompanied by the on-board entertainment of NCAF's own Fenella the Rottweiler. The night will be rounded off at the riverside Victoria Arms, before punting (slowly) home.

Whether you are a new face to this subject or an established member, whether based in academia or industry, whether interested in signal processing, biomedical engineering, or more general 'Information Engineering' discussions, we look forward to welcoming you to Oxford.

David Clifton and Mark Ebden
University of Oxford

INSIDE

BACK PAGE

Beethoven, Bayes and beer in Birmingham

Gene micro-array analysis can identify the presence or absence of many thousands of genes.

Machine learning was the theme for the winter NCAF meeting, held at Aston University. General papers were supplemented by the first Student Forum, in which several PhD students described their current research.

Ian Nabney (Aston University) started the meeting talking about the hidden connections that may be found in the wealth of data that is now almost routinely collected. Questions like 'do I have a particular disease?' cannot be solved by an algorithm, but experts in the field generally have the experience to answer them. In this example, the experts are doctors and the experience comprises input data (the many patients and their clinical symptoms) and output data (whether or not the patient actually had the disease). A good model must be able to generalise as it is usually impossible to list all combinations of input values (even something as simple as character recognition has about 10^{190} input states). Models must also be able to cope with noise, uncertainty and the presence of outliers. At this point, Ian played a recording of himself playing Beethoven's Opus 111. This piece is considered an outlier in terms of musical syncopation, being over 100 years ahead of its time. Looking to the future, problems will get larger (more data), there will be more methods to analyse them, and the data will be in different forms (e.g., images or based upon natural language).

The second talk, from Mike Osborne (Oxford University), continued the theme and considered how Gaussian processes (GPs) can analyse online data. GPs permit the use of a Bayesian framework, but also provide error bars that allow only the most useful GPs to contribute data to the model. The approach has been tested using weather data collected from four sensors situated in the Solent, available on the Internet. Data missing from one sensor can be imputed from the remaining three.

Gene micro-array analysis

Mike Tipping (Vector Anomaly) also considered how Bayesian learning can lead to meaningful models. It is more important to have a meaningful model than an accurate one. A Bayesian methodology implements the concept of Ockham's Razor, which can be loosely paraphrased as the simplest solution is usually the best. Gene micro-array analysis can identify the presence or absence of many thousands of genes. With relatively few patient samples, it is nearly always possible to develop an accurate model, but with no underlying science. However, a sparse Bayesian model can, just enough to be accurate but without over-complication. See Mike's description at www.relevancevector.com for more information.

Ata Kaban (Birmingham University) gave a talk describing the combination of data modelling and visualisation. The eye is excellent at visual analysis, and can identify anomalous data values without having to define an algorithm. She used colour to draw the eye to key values. These methods were applied to generative topographic maps of web navigation sequences. The objective is to help future navigation by predicting the next web page to be viewed.

David Clifton (Oxford University) described novelty detection with extreme value statistics. A problem exists when trying to develop a model to analyse faults in something reliable like an aircraft engine. There is a surfeit of normal data, and too little abnormal data. David divided his data into sections, and then only considered the extreme values within each section. Novelty thresholds were set, to define the boundary between normal and abnormal behaviour. The methods have been applied to the analysis of aerospace gas-turbine engines and have provided indications of potential engine faults. The method can be modified to analyse inliers, not just outliers.

The first part of the PhD forum was next, with two speakers. Thomas Bermudez (Aston University) spoke on the detection of brain malfunction from EEG and ECG data. Martin Schroeder (Aston University) described his exploration of geochemical data with missing values.

The presentations ended with 'Puzzle Corner' and the traditional method of acting out the solution. Apparently, all gregarious academics drink copiously. This was put to the test during the social event, a trip on a Beer Bus to the Black Country. This outing was made all the more enjoyable by the fiendish quiz compiled by Vicky Bond.

Optical astronomy

Activity resumed promptly on the morning of the second day after the intellectual (and other) challenges of the social event. Our first speaker, Peter Tino (Birmingham University), made a dramatic entrance in full cycling gear, having dashed across the city in heavy traffic. His talk was on the development of a probabilistic model for two time series representing delayed, irregularly sampled and noisy versions of the same underlying pattern. This has been applied to a problem in optical astronomy, where the delay in the signal from a quasar can be used to estimate the mass of a large intervening galaxy (based on gravitational lensing). With a large series of systematic cross-validation experiments he showed how he could beat state-of-the-art methods used by astronomers because he was working from a different mindset.

Dan Cornford (Aston University) gave a fascinating overview of different ways in which physical and statistical models can be combined. Currently one can either build deterministic physical models (often based on differential equations) which are simulated numerically or statistical models (based on a simple combination of variables and noise) which are used in a Bayesian inference framework. It is clear that for many large-scale models, particularly in an environmental context, neither of these is adequate: physical models give limited and poor quality information about variations, while statistical models are too limited and simplistic. Dan proposes developing a framework of stochastic simulation where realistic complex physically-based priors are combined in a probabilistic model with advanced inference methods to assimilate data and make predictions.

The final talk in the morning was from Richard Watson (Southampton University) who is developing new types of genetic algorithms which are designed to

How to avoid making the headlines – simple and secure data encryption

NCAF loses membership details

A headline that none of us want to see

Not all publicity is good publicity, and in the last year, there have been many high profile cases of personal data being 'mislaid' by government departments and private companies. These affect public confidence, bring possible legal actions against the organisation and leave the individuals concerned open to identify theft.

Data losses happen in a variety of ways, including resale of unwiped hard disks or posting of un-encrypted CDs. Some, like theft of a laptop or losing a USB memory stick, could happen to anyone. If you have data which you would not publish on the Internet for anyone to see, then you should be encrypting it. I would recommend encrypting all your documents and data which you store. Although mathematically complicated, encryption is simply the scrambling of information so that it can only be read using a key. 'The Code Book' by Simon Singh is a very readable introduction to the subject. One important thing to note is that password protecting a laptop, a database or a spreadsheet does not really encrypt your data. These methods can all be cracked in minutes and only give a false sense of security.

How to encrypt your files

The best way is to set up what is known as an 'encrypted volume'. This is typically a large file which can be used just like an extra disk. In my opinion, the best product for this is the open source TrueCrypt, using encryption approved by the US government for top secret information.

You can download TrueCrypt from http://www.truecrypt.org/. It is easy to use and the expanded version of this article at http://www.the-data-mine.co.uk/encryptingdata has links to online guides. As always, you should make a backup before proceeding.

Setting up your encrypted volume is simple; just stick to the default values for formatting and encryption. Make your virtual disk big, as almost everything you store will end up in it. Enter a password – and do not forget it as there is no other way to get access to your data! Mount the container file from TrueCrypt, and move your data files across to the drive. You are now secure!

How to encrypt data for distribution on CD

There are many factors other than encryption to consider when shipping databases containing confidential information. For example, what will the recipient do with the data? Will they keep it secure? The brief guide below shows you how to securely encrypt data in a way which means that however many copies you lose in the post, there is no security risk.

First, you create a TrueCrypt volume containing the data, copy it to a CD, and post it off. When (or should this be if) it arrives, and is inserted into a PC, the user is prompted for the password or phrase. You should supply this via another channel such as a phone call or a registered letter via a different courier. The encrypted files can now be read just like normal files. Any adversary would need to intercept both the CD and the password or phrase. If you are paranoid, you can use some form of tamper evident packaging for the CD so the intended recipient can be sure that it has not been copied during its journey.

Tips

- TrueCrypt can start automatically when your machine boots so dismount the encrypted drives when you hibernate or shut down.
- On Windows, you can move the location of your 'My Documents' folder to the encrypted disk by right-clicking on it.
- USB memory sticks can be encrypted and automatically start TrueCrypt when inserted, allowing you to transfer data safely.
- For extra security, run a program to clean up your disk once you have encrypted the files.

Dr Andy Pryke
The Data Mine Ltd

work on problems where a divide-and-conquer approach is needed. Conventional genetic algorithms fail, because gradual improvement methods cannot find the optimal solution.

There followed the second half of the PhD forum, with three speakers. James Hensman (Sheffield University) is using Gaussian processes to locate cracks in complex structures from acoustic emissions; Jack Raymond (Aston University) is applying analytical methods from statistical physics to understand attacks on fully-connected networks; and Steve Dalton (Aston University) is developing genetic programming algorithms to perform cluster analysis. At the end of the meeting there was a vote by the audience, carried out using single transferable votes. After a hot contest and much counting, the first prize was awarded to Thomas Bermudez and the second prize to James Hensman.

Another successful meeting ended. Thanks are due to Vicky Bond for being the local organiser. NCAF is grateful to Springer, publisher of the NCAF Journal, for supplying the prizes for the PhD forum.

Ian Nabney (Aston University)
Nick Granville (Smith & Nephew Technical Services Group)

PUZZLE CORNER

Number 38

Following the infamous cocktail party thrown by the now impecunious Fontainebleau Plantagenet VII, Lisa was invited to partake in her celebrated Egg Hiding game. One person, the Coopier, was required to secrete a valuable Phabberjay Egg in one of 10 boxes numbered from 1 to 10. The second player, the Foxy One, would then gamble on being able to find the egg, using the following procedure.

Starting with the lowest numbered box, the Foxy One would decide whether to gamble on that box, or not. If gambling, he would pay the wager (1000 Euros) and could have whatever was in the box, and then that particular game would end. Obviously if the egg is elsewhere he would just have lost his wager. If he chooses not to gamble on that box, then it is opened and, if the egg is present, then the game ends (but at least he has not lost the wager). If it is empty, then play proceeds to the next numbered box.

Obviously, if the Foxy One chooses not to gamble on the first 9 boxes, and the egg does not appear, then he can 'gamble' on the last box safe in the knowledge he will soon have his very own Phabberjay Egg. It is also obvious that placing the egg in box 10 is a risky business for the Coopier – or is it?

If the value of a Phabberjay Egg is 2000 Euros, what would be a fair premium (per game) for the Foxy One to pay the Coopier for the privilege of playing?

The answer will be given at the next NCAF meeting (18–19 June 2008, Oxford University).

Fenella the Rottweiler

ncaf Natural Computing Applications Forum

COMMITTEE NOTES

MEMBERS' NEWS AND VIEWS

Deadline for contributions for the next edition – 1 July 2008.
Please send to Managing Editor – Nick Granville, e-mail:Nick.Granville@smith-nephew.com

Edited and Produced by:
Forum Public Relations
Westgate House, Old Ivy Lane
West End, Southampton
Hampshire SO30 3RX
Tel/Fax: 023 8013 9805
e-mail: info@forum-pr.co.uk

NEXT EDITION
Review of Oxford meeting
Preview of the Autumn meeting

Neuroinformatics or whatever happened to neural networks

Technology keeps breaking new ground, but there is a surprising level of circularity to many developments. Neural networks could well be a case in point.

NCAF members will recall that until recently 'NCAF' stood for the 'Neural Computing Applications Forum' and had a strong bias towards neural networks. Some older members will recall that neural networks were inspired by extremely simplistic computer models of the brain which serendipitously performed well at empirical modelling tasks applied in commerce and engineering. NCAF and the mainstream artificial neural network community have subsequently moved in the direction of statistically rigorous methods and the phrase 'principled approach' has become almost *de rigeur*.

What has happened to real neural network research in the meantime? Are there any activities relevant to NCAF?

Neurobiological research involves simulation and modelling, in-vivo measurements and in-vitro measurements all aimed at advancing the understanding of the structure and function of the brain. Experimental neuroscience generates vast quantities of data which can be difficult to share between groups of researchers due to differences in experimental protocol and the difficulty of indexing large quantities of relatively unstructured data. It is not surprising that searching databases is tricky – someone recording neuron spike trains may observe an interesting feature in the data as a result of a particular chemical or physical stimulus. How does one check whether anyone else has observed a similar effect? How might one index what amount to wiggles in very long traces? Which experiments are likely to be representative of a particular set of conditions? Which analysis tools are appropriate for a particular type of experimental data? How can computer models be validated and compared?


Inscentinel

All-year round bee supply

In 2007 the Medical Research Council provided funding for the UK to join the International Neuroinformatics Co-ordination Facility (INCF, www.incf.org). This organisation, set up by OECD, moved into new offices in Sweden in 2007. Its aim is to develop and apply advanced tools and approaches essential for a major advancement in understanding the structure and function of the brain. The UK node is co-ordinated by Professor David Willshaw and based at the School of Informatics within the University of Edinburgh.

The UK node has made contact with several of the country's leading research groups which are active in this field. One collaborative project may be familiar to NCAF members. It is CARMEN an e-Science project (www.carmen.org.uk) supported by the Engineering and Physical Sciences Research Council. The objective of the project is to create a virtual laboratory in which data on neuronal activity (electrical and optical measures) can be shared, stored, manipulated and modelled. The project currently involves a consortium of 20 academic investigators from 11 universities as well as commercial associates, but will expand in the future. Several statistical or neural network methods used in CARMEN can be traced back to the e-Science project DAME which aimed to manage health monitoring data and, even earlier, to the 'Neural Networks – The Key Questions' projects at York and Oxford Universities.

It is interesting to see the circularity by which neurobiology originally inspired artificial neural networks and now the statistically rigorous neural network methods are being used as neuroinformatics tools.

CARMEN is far from being a one off. Apiculture is another topic where there has been a circular sequence of technology development, loosely involving neural networks. Complex behaviour observed in insects, despite the relative simplicity of their brains, inspired IT developments in the areas of group behaviour and pattern recognition. More recently real bees have been used for pattern recognition in the Inscentinel product (www.inscentinel.com) which uses bees to detect explosives, drugs and fruit. In pursuit of this, the BBSRC Rothampstead Laboratory, the developers of Inscentinel, use some simple but elegant image recognition software to detect the Pavlovian feeding reflex in trained bees. A recent study, reported in New Scientist (26 January 2008), described how bees select a new nest site – by forming a swarm consensus on the best of several candidate sites each visited by different small groups of scout bees. The research has discovered that the 'voting' mechanism used by the 20,000 strong swarm appears near optimal and very similar to the way in which strongly excited neurons in the brain 'recruit' other neurons to reinforce a newly learned response.

It is too early to tell whether there will be a breakthrough in understanding of the brain and whether this will inform a new generation of biologically inspired computing. However, the author's view is that a biologically inspired computing paradigm will overcome the noise and communication limitations of deep sub-micron computing devices.

Peter Cowley
Quarndon Cognition Ltd

DIARY DATES 2008

16–18 July – NCPW11, the 11th Neural Computation and Psychology Workshop, Oxford, UK. http://psyweb.psy.ox.ac.uk/babylab/NCPW/index.html

3–6 September – ICANN 2008, 18th International Conference on Artificial Neural Networks, Prague, Czech Republic. http://www.icann2008.org/

September meeting – NCAF meeting (theme to be confirmed) at Sheffield University.
For information, email enquiries@ncaf.org.uk or telephone +44 (0)1332 246989.