Networks

Intelligent Diagnostics

NCAF Summer Meeting, University of Sunderland 9–10 June 2004

CAS was established in 1996 with the goal of conducting applied research and development work in a range of adaptive computing techniques: neural networks, genetic algorithms, neuro-fuzzy systems, case-based reasoning, and other paradigms based on natural problem solving and reasoning.

NCAF is returning to Sunderland after a gap of some six years for the summer meeting. Previous meetings hosted at Sunderland in 1996 and 1998 have been great successes and both of these focused very much on the applications of neural computing and related techniques. Industrial participation in these events was very good.

This summer's meeting should continue that trend. The theme, Intelligent Diagnostics, is one of the key areas of activity of the Centre for Adaptive Systems (CAS) at the University of Sunderland. CAS was established in 1996 with the goal of conducting applied research and development work in a range of adaptive computing techniques: neural networks, genetic algorithms, neuro-fuzzy systems, case-based reasoning, and other paradigms based on natural problem solving and reasoning. The Centre has, since its inception, been very active in the areas of condition monitoring, predictive maintenance, fault diagnosis, prognosis, signal processing, and pattern recognition. Almost all of this work has been applied to real problems with industrial partners. Some industrial partners exploited the resulting patents and now have products in the market place.

The Centre has been a partner in more than a dozen large European research projects, collaborating with key industrial partners such as Kone Elevators (Finland), Rockwell Automation (UK and The Netherlands), NPower, Monition International, Diagnostic Solutions, Interface Condition Monitoring, and many others.

The programme for the June NCAF meeting is in two parts: the first day has the theme of Intelligent Diagnostics and will include three presentations from CAS research staff on various current projects in the field of condition monitoring and predictive maintenance. Dale Addison will present on his work in the MINICON project, which has developed an intelligent diagnostic system with two application areas – embedded diagnostics in elevators (a major maintenance problem throughout Europe), and machine tool diagnostics (a notoriously difficult application). Dr. Adam Adgar will present on his

work on the SENSOIL project, which focuses on the intelligent analysis of data coming from new sensors (designed by other partners in the project) giving data on oil quality and condition. Professor Bob Mattheys will describe the work of the Intelligent Systems Solutions team – a regionally-funded project which is working with North East SMEs to demonstrate the benefits and potential of intelligent systems in a range of applications. Bob will describe a specific diagnostic system developed for a local engineering company.

The second day of the meeting will, as usual, move away from the theme and include presentations across a range of topics and application areas. Dr. Malcolm Farrow, Co-Director of CAS, will describe his work on medical data analysis using Bayesian Networks, and Dr. Giles Oatley will present his work – also using Bayesian techniques – on the analysis of crime data for prediction of repeat offences. These will be complemented by a range of other talks.

The social event, always an important part of the NCAF meeting, will be a return to Beamish Open Air Museum, set in over 300 acres of beautiful countryside, vividly recreates life in the North of England in the early 1800s and 1900s. Winner of both the British Museum of the Year and European Museum of the Year Awards, it demonstrates the recent history of the region in a "living" way and provides entertainment and education for visitors of all ages and interests. NCAF delegates will take the tram from the entrance and enter the village, discovering a typical North East village, circa 1913, complete with the famous Sun Inn, a pub that some members of the NCAF "family" will remember very well.

We look forward very much to seeing you all at Sunderland in June, with a friendly North East welcome, an excellent technical programme, and a very enjoyable social event at which we can catch up with old friends and colleagues.

Professor John MacIntyre
University of Sunderland

Discover who

Human speech has several levels. The simplest is the phonetic level; phonemes combine to form syllables; then there is the lexical level that combines to give syntax. Ferret works by analysing trigrams, or three adjacent words. Only 5% of trigrams will occur in each of two articles written by different authors, even in articles on similar topics.

N CAF held its winter meeting at Bath, with the theme of 'Fundamentals of Natural Computing' and the enigmatic title of 'Back$_6$'. The theme was a direct result of a response to the recent questionnaire suggesting that a more introductory level meeting would be useful. This resulted in having fewer but longer talks than normal on the first day, allowing each speaker more time to cover his or her area of natural computing.

First up was Andy Webb (QinetiQ) who gave a tutorial on the statistical approaches to pattern recognition. The data can be collected from many sensors and be diverse (binary, categorical, or continuous; it can also be subjective). It could also be highly dimensional and have missing values in a large dataset. The generic approach to analysing the data might involve extracting features from each set of sensor data, and then classifying the set of features from all sensors in order to make a decision. The classification can either be supervised or unsupervised, but if the latter, how many classes do you choose? A further problem is that the different dimensions in the data need not be independent. Andy described one dataset of 500 samples, with six categorical dimensions (of 8, 7, 7, 3, 3 and 2 classes making a total of 7,056 bins). A nearest neighbour approach says that for a given sample x, find k nearest neighbours, and then classify x by majority voting of the neighbours. It is also possible to have non-linear discrimination by changing the majority voting to some other method that incorporated the differences (or distance) between the nearest neighbours. He ended by saying that although the classification performance could be described by some mathematical function, determining whether it was fit for purpose was a better measure of effectiveness.

from the inputs. Steve explained that the optimal size of a network was the minimum size that explained the majority (or sufficient) of the data. Regularisation is a method in which over-complicated solutions can be penalised and discouraged. This will lessen the chance of over-fitting, i.e. fitting the noise in the data and reducing the ability of the network to perform with unseen data. There are two main measures of performance. Sensitivity and specificity are important for classification networks, whereas the sum of least squared errors is important for regression networks. Finally, do not be afraid to use more than one network on your data. Committees of networks, each making one prediction and then voting to determine the best prediction, nearly always perform better than just one network.

Tom Khabaza (SPSS) then described the general principles of data mining (DM). Whereas domain knowledge is desirable but not essential when using a neural network to analyse your data, it is essential if you intend to start data mining. You also require sufficient data that are likely to contain the patterns that you are looking for. DM is also a repetitive process, unlike neural networks (which are generally optimised and then applied to more data). Tom gave the analogy that traditional statistics is about hypothesis testing whereas DM is about hypothesis generation. A good DM tool will be able to identify clusters of similar data, or associations between subsets of the data. Clustering can be unsupervised and can be used to find patterns in input data without corresponding output data. Unlike the output from an NN, DM rules usually 'read well' and can (and should) make sense. Finally, Tom pointed out that few DM projects fail. Snippets of useful information are usually found, even if the main goal is missed.

Segway Human Transporter

One technique

Steve Roberts (University of Oxford) then built on the first tutorial by considering just one technique of implementing pattern recognition and prediction – neural networks (NN). NN's are probably the most established natural computing technique, but they do not provide 'turnkey' solutions. Key issues are the type and size of the network. Multi-layer perceptrons (MLP) and radial basis functions (RBF) networks are the most common. MLP's work by calculating an output from a set of input values (each linearly weighted and squashed into a small range (typically 0→1) by an activation function). RBF's work by calculating the weighted sum of the distance of a set of input values from a set of prototypes, and therefore, are more often used to classify a set of inputs. MLP's, on the other hand, can be used either to categorise a state into 1 of n classes, or to calculate some variable

Evolutionary computing

The final tutorial of the day was from Xin Yao (University of Birmingham) who spoke about evolutionary computing and genetic algorithms (GA). This talk contrasted very nicely with the previous talk in that domain knowledge is not a requirement for the successful implementation of a GA (although it can help). GA's require genes and these can crossover or mutate, just like their biological counterparts. Crossover is the method used to find the optimal (or near optimal) solution. Mutation is the method used to ensure that the entire solution space is searched. Mutation must occur infrequently because it usually introduces regressive genes, but it is the only method of introducing new genetic information into the gene pool. Finally, Xin described evolutionary learning. This could describe a rule-based system in which the rules evolve, and could be fuzzy, or an evolutionary

did it in Bath

neural network in which the connections between the nodes come and go, rather than just the weights changing.

Caroline Lyon (University of Hertfordshire) described the Ferret Copy Detector, a software tool that can analyse text for plagiarism. Human speech has several levels. The simplest is the phonetic level; phonemes combine to form syllables; then the lexical level combines to give syntax. Speech recognition generally works using hidden Markov models, whereas Ferret works by analysing trigrams, or three adjacent words. Only 5% of trigrams will occur in each of two articles written by different authors, even in articles on similar topics. However, this percentage can increase to 15% when plagiarism occurs. Caroline went on to give an impressive live demonstration of the software during Day 2.

Graham Hesketh then introduced a brief workshop session in which the audience could describe some of their own problems (there weren't many) before putting on his puzzle corner hat. Graham admitted that initially he had been unable to solve the problem, even though his 15 year-old son had. Let that be an inspiration for all you NCAF members that do not even attempt the problem. The social event was a murder mystery acted out over a sumptuous meal in the Assembly Rooms. With characters like Binky and Squealer (and they were members of the audience), the evening was a great success. Ex-Inspector Adderley's team was deemed to have been nearest to identifying the murderer. But even they were wrong, and shouts of 'Unfair, he's a professional' went unheeded by the winning team as they collected their prize.

Visualisation methodologies

Carren Holden (BAE Systems) started the second day describing visualisation methodologies in aircraft design. Problems occur when there are a large number of variables (8 or more dimensions is not uncommon), each valuation is time consuming, and the optimum solution is hard to find. Data in two dimensions is easy to visualise from the colour in a 2-D plot, called a tile. Data in four dimensions can be visualised by having a 2-D array of tiles, called a table. The colour plots gradually change from tile to tile, and it is not too difficult to identify the optimum solution. Data in six dimensions can be visualised by having a 2-D array of tables, although it is now more difficult to use one's eyes to determine the best solution. Carren also reduced the number of dimensions in complex problems to make the visualisation easier. In a very large 22-D problem, reducing the dimensions to five gave the best solution. The 22-D solution was too hard to search. Patent GB0323848.2 (filed 10/10/03) gives more information.

After a demonstration of the Segway Human Transporter (see www.segway.com for more details), Nigel Shadbolt (University of Southampton) talked about Engineering Webs of Knowledge. There are about 10^8 users of the web, and about 10^{10} indexed resources on Google. However, Google searches rarely take longer than a few seconds. Nigel described the AKT or Advanced Knowledge Technologies project which aims to implement an ontology (a shared and controlled vocabulary and grammar) so that searching the web for knowledge remains just as quick in the years to come. There are three levels of data. Data is the lowest, and is a collection of facts; information is higher and comprises data in context; the highest level is knowledge, which is information that enables decision-making. Sharing knowledge often has sociological problems, but is only possible if people want to do it and the knowledge is presented in the right format (e.g. using a shared vocabulary). There is now an ontology standard (OWL). The challenges lie in ensuring that the ontology evolves while remaining in a controlled environment, ensuring the integrity of any knowledge on the web, and finally ensuring that the users trust the provenance of the knowledge that they find.

Tom Musicka (University of Newcastle) returned to the topic of data analysis, and considered the case of data generated from a hot strip steel-rolling machine. Data were in ample supply, but had drift due to the wear of some machine parts superimposed. Even after the removal of outliers and the use of as much process knowledge as possible, the neural networks still gave poor predictions if the drift was ignored. As mentioned above, a committee of networks gave better results. The final talk was given by Meghana Ransing (University of Swansea) and compared the performance of neural networks with other less 'black box' methods such as regression analysis or the use of expert knowledge. Human experts can only cope with linear or quadratic relationships, and therefore cannot determine some of the complex non-linear relationships revealed by an optimised neural network. However, with many input nodes and relatively little data, most NN's do not perform well. The advice was to use as much domain knowledge as possible.

Despite being themed around the fundamentals of natural computing, most attendees admitted learning something new from the meeting. It is a theme that NCAF should return to every few years to ensure that newcomers to the field have a good understanding of the different methods.

Thanks must go to 'Bert Bullen for his very efficient organisation of the meeting, and to BAE Systems for their sponsorship of the meeting.

Nick Granville
Smith & Nephew Research Centre

COMMITTEE NEWS

At the recent AGM in January at Bath University, two stalwart members of the Committee stood down after many years of service. Rick Adderley, a long-standing member and outgoing Editor of *Networks*, retired from the police force to set up his own company, AE Solutions. We wish him every success in his new venture. Mark Cheeseman, the outgoing Secretary, has endured a challenging 3-year term in the post. I cannot overstate how much support Mark has provided to NCAF and myself. His determination, enthusiasm and professionalism were great assets to the organisation, and we are very grateful for the solid foundation he has created.

Three existing Committee members stood again and were re-elected: Graham Hesketh, Ian Nabney and Richard Everson. In addition, two new members were elected: Armin Stranjak (Rolls-Royce) and Nick Granville (Smith & Nephew).

At the Committee meeting on the 4th February, the NCAF Officers for 2004 were elected: Chairman (Graham Hesketh), Secretary (Armin Stranjak), and Treasurer (Ian Nabney). Also, Nick Granville accepted the post of Editor of *Networks*.

Graham Hesketh
Chairman

A message from your new editor

As Graham Hesketh has mentioned above, Rick Adderley has resigned from the Committee and as Editor of *Networks*. I accepted the job when Graham caught me off-guard. I hope that I can maintain the high standards that Rick set.

One thing that always strikes me about the NCAF meetings, apart from their friendliness, is how interested everyone is in each others work, and in the analysis methods that they are currently using. It makes a pleasant change from some of the other scientific or medical conferences that I attend, where confidentiality is more important. Of course, NCAF is fortunate in that few of its industrial members are competitors of other members. I would like *Networks* to capture the spirit of the meetings. I am also keen that the current decline in the number of industrial members of NCAF is halted, and even reversed.

All academic institutions carry out research, and are eager to publicise their results. They are also usually keen to improve their links with industry by carrying out more specialised projects funded by industry. But industrial members may not know what each university specialises in, particularly if they are new to the area of natural computing. Therefore, I intend to ask representatives from academia to write a few hundred words describing what they do, and how they could help new members of NCAF. Particular areas of interest and short courses could be described, as could potential topics for collaborative research. Almost anything can be mentioned; I want to capture the essence of each department.

With only three issues of *Networks* each year, it may take me a long time to work round all institutions. So if you have anything that you want to publicise, please volunteer your copy now.

Nick Granville
Smith & Nephew Research Centre

DIARY DATES 2004

9–10 June – NCAF Meeting on Intelligent Fault Diagnostics at the University of Sunderland. For information, e-mail enquiries@ncaf.org.uk or telephone +44 (0)1332 246989.

26–30 June – GECCO'2004: The Genetic and Evolutionary Computation Conference, Seattle, Washington, USA. http://www.isgec.org/gecco-2004/

12–14 July – CIMCA'2004: International Conference on Computational Intelligence for Modelling Control and Automation, Gold Coast, Australia. http://www.ise.canberra.edu.au/conferences/cimca04/

13–17 July – SAB'04: The Eighth International Conference on the SIMULATION OF ADAPTIVE BEHAVIOR, Los Angeles, USA. http://www.isab.org/sab04/

23–25 August – COMADEM 2004: The 17th International Congress & Exhibition on Condition Monitoring and Diagnostic Engineering Management, Robinson College, Cambridge, UK. http://www.comadem.com/2004.htm

29 August – 1 September – BICS'2004: Brain Inspired Cognitive Systems, University of Stirling, Scotland. http://www.icsc-naiso.org/conferences/bics2004/program.html

15–16 September – NCAF Meeting (theme TBD) at the University of Swansea. For information, e-mail enquiries@ncaf.org.uk or telephone +44 (0)1332 246989.

MEMBERS' NEWS AND VIEWS

Deadline for contributions for the next edition – 7 July 2004. Please send to Managing Editor – Nick Granville, e-mail: Nick.Granville@smith-nephew.com