IMA Seminar Series 2008/2009

Contact:

Peer-Olaf Siebers
Phone: 01158466526
Email: pos@cs.nott.ac.uk

23 June 2009, 2-3pm, Room C60

Seminar by Daniele Soria: Application of Affinity Propagation on Breast Cancer Data Sets

The Affinity Propagation algorithm is a novel clustering method proposed by Frey and Dueck in 2007 [1], which combines advantages of both affinity-based clustering (like Hierarchical algorithms) and model-based clustering (like Expectation-Maximisation). The method takes as input measures of similarity between pairs of data points and real-valued messages are exchanged between data points until a set of centres (called exemplars) and corresponding clusters gradually emerge. The Affinity Propagation algorithm also provides a procedure to determine the number of clusters to be considered.

In this talk, the approach proposed by Frey and Dueck will be presented and results coming from the application of the technique to breast cancer data sets will be reported and compared with the original ones. Results from Affinity Propagation are consistent with the results already obtained having the advantage of providing an indication about the number of clusters to consider in the analysis. Moreover, results also provide novel insights, with respect to conventional approaches.

[1] B. Frey and D. Dueck. Clustering by passing messages between data points. Science, 315:9726, 2007.

16 June 2009, 2-3pm, Room C60

Seminar by Jamie Twycross: Executable Modelling for Plant Systems Biology

My research explores the application of executable models to systems biology. In collaboration with biologists, I am developing a number of multiscale executable models investigating the transport of plant hormones, and the effects of these hormones on root phenotypes. Executable models differ from mathematical models in that an executable model defines a procedure which is applied to transform the modelled system from one state to the next. The model is simulated, rather than solved, to determine the time evolution of the modelled system. I am using a stochastic P system framework to describe my models. Models can contain many hundreds or thousands of compartments, molecular species, and reactions, corresponding to complex multicellular tissues, signalling and transportation pathways. Specification of such models can be eased though the use of modularity, which allows commonly repeated sets of rules (modules) to be specified only once. Using P systems, executable models of biological systems can be constructed without the need to write down complex systems of differential equations, permitting preliminary studies to be performed in silico and reducing time spent in the laboratory. I will present a general introduction to executable models and P systems for plant systems biology, overview the software I have developed, and outline some of my recent work in developing executable models of plant growth.

09 June 2009, 2-3pm, Room C60

Tutorial by Peer-Olaf Siebers: Simulation with AnyLogic

The focus of my tutorial will be on AnyLogic, a multi-paradigm simulation software that we use in our research group to conduct some research on complex systems modelling and simulation. In AnyLogic simulation models are designed in an object oriented fashion. The software supports all common simulation methodologies (system dynamics, process centric, individual based) and can be used to capture the complexity and heterogeneity of business, economic, social and biological systems to any desired level of detail.

I will give a demonstration on how to develop a simulation model with AnyLogic and how to run experiments with it.

02 June 2009, 2-3pm, Room C60

Seminar by Navapat Khantonthong: Introduction to Neural Networks

Neural Network is the representation of brain's learning approach. This brain operates as multiprocessor and has excellent interlinked. Neural Network also can be represented as "Parallel distributed processing" planning. It is utilised in the computer applications for solving the complicated problems. There are many benefits from Neural Network such as no requirements for specifying the relevant factors, an unsophisticated model which has many factors for performance, a straightforward model, fault tolerance and an innate synchronous.

In this presentation, there are two types of Neural Network, which are implemented for investigation. Kohonen's feature map is also called as Kohonen Self-Organising Neural Network, and is an unsupervised learning. It has the efficiency of learning with no guidance. Furthermore, Back Propagation Neural Network is a type of Multilayer Perceptron in FeedForward architecture and is a supervised learning. It has learning caused by the many performances of a recommend set of training sample to Multilayer Perceptron. As a result, these Neural Networks are applied for solving the example problems such as Matrices, Travelling Salesman and XOR. They will also adapt for the research about the Biometric Thrill and Arousal Detection system in the future work.

26 May 2009, 2-3pm, Room C60

Seminar by Xiao Ying Wang (Sally): Methods of Interpretation of a Non-stationary Fuzzy System for the Treatment of Breast Cancer

Recommending appropriate follow-up treatment options to patients after diagnosis and primary (usually surgical) treatment of breast cancer is a complex decision making problem. Often, the decision is reached by consensus from a multi-disciplinary team of oncologists, radiologists, surgeons and pathologists. Non-stationary fuzzy sets have been proposed as a mechanism to represent and reason with the knowledge of such multiple experts.

In this talk, I will briefly present the creation of a non-stationary fuzzy inference system to provide decision support in this context, and examine a number of alternative methods for interpreting the output of such a non-stationary inference system. The alternative interpretation methodologies and the experiments carried out to compare these methods are displayed. Results shows that using majority voting ensemble decision making from a non-stationary fuzzy system improves accuracy of the decision making.

IMA Seminars on 19 May 2009, 12-1pm, Room C60

Seminar by Francisco Bonachela Capdevila (Katholieke Universiteit Leuven Campus Kortrijk, Belgium): Application of Clustering Techniques in Gene Prioritization Using Endeavour

Very common diseases in our society, such as cardiovascular problems, diabetes, schizophrenia and some types of cancer go beyond Mendel's model of inheritance.

Application of classic methods like linkage analysis and positional cloning to these diseases can become a very laborious task since several areas of the genome can be identified, containing too many genes to be cloned and investigated in the laboratory. A reliable approach to deal with these complex diseases is the candidate gene approach, where genes are ranked according to different parameters and then the best classified ones are cloned and investigated in laboratories. Endeavour is a software tool developed by SISTA group in University of Leuven that takes advantage of this strategy. It uses several heterogeneous data sources to classify a complete set of candidate genes according to their similarity with a selected group of training genes.

Though the original experiments made by SISTA group have been very positive there is still a long way to go since there are some diseases where Endeavour performance is not that good. The reason for this lack of performance is not clear. One example is leukaemia, with a training set of more than 100 genes. Probably, such a extensive training set is not homogeneous enough to establish a clear pattern to precisely classify the genes of the candidate set. Our strategy to cope with this drawback is unsupervised classification.

In this talk we report on our current work on the application of clustering techniques to the first phase of the prioritization in order to obtain more homogeneous training sets and therefore more accurate results.

12 May 2009, 2-3pm, Room C60

Seminar by Cesar Caballero: Single-trial BOLD fMRI

Functional Magnetic Resonance imaging is employed to map in space and time the brain's response whereas a subject is performing some kind of cognitive task. Traditionally, fMRI data analysis techniques have employed prior information about the experiment, or paradigm, in order to offer statistical assessment of the activations. I will present a novel method which avoids this assumption and releases the analysis of timing constraints. The method is based on the deconvolution of the fMRI time series and the definition of a T-statistic thresholded by the False Discovery Rate procedure. Having tested its performance in a visually-cued and self-paced motor paradigm, our results have demonstrated that with this method it is possible to detect cortical responses to single trials without prior information of the paradigm. The study also has direct implications in the study of spontaneous brain activity at rest.

Seminar by Shabbar Naqvi: An investigation into the use of Fourier Transform Infrared Micro-Spectroscopy for the automation of Breast Cancer Grading

Breast Cancer, which has the highest incidence rate in women, is also the most common cancer in UK. Medical Prognosis plays a vital role in approximating the re-occurrence of disease and survival of patients. Nottingham Prognostic Index (NPI) is widely used in the world for the prognosis of breast cancer. It consists of three parameters namely tumour size, lymph node status and tumour grade. Out of these, grade is an crucial parameter. Nottingham Grading System (NGS) is used to find the grade in NPI. In this talk, I would present the background of NPI and NGS and efforts made in the past for the automation of NGS and potential of Fourier Transform Infrared Micro-Spectroscopy (FTIR) for this purpose.

5 May 2009, 2-3pm, Dearing Building, Room C42

Seminar by Duc Thang Ho: A Fuzzy Approach for the 2007 CIG Simulated Car Racing Competition

In my presentation I will describe the techniques that have been used by the winning entry of the 2007 IEEE Congress on Evolutionary Computation (CEC2007) and the CIG2007 car racing competitions. The challenge is to race against an opponent around a track, trying to get as many points as possible. Previous research on similar problems are mostly based on either state-based or action-based controller architectures trained with machine learning techniques. I will present a hybrid controller architecture, combining both the advantages of the existing architectures. The main component of the controller is designed as fuzzy systems whose membership functions are changeable according to the context. Finally, the competition results are given.

28 April 2009, 2-3pm, Room C60

Seminar by Noor Azizah KS Mohamadali: Evaluation Studies in Health Informatics and a Proposed Integrated Model of Technology Acceptance for Health Information Systems

The evaluation of software is an area of research in the health informatics field which is being studied quite extensively. Evaluation studies are necessary in order to ensure that systems adequately meet the requirements and information needs of the users and organization as a whole. Evaluating what are the critical success factors of implementation of new information systems is crucial because there is evidence to show that quite large numbers of new systems have failed to be implemented. There is also evidence that the same system implemented in two different departments resulted in two different outcomes; in one department it was accepted by the users, while in the other it was rejected. It is believed that success of systems operating in particular environments largely depend on user acceptance of the system. Identifying those critical factors that influence user acceptance of new technology is an interesting and challenging problem which needs investigation.

In this presentation, I will give an overview of evaluation studies in health informatics, analyse existing information system theories on user acceptance, analyse existing work on user acceptance of technology in healthcare, and discuss a proposed integrated model of technology acceptance which incorporate those critical success factors. I will also outline my future work.

Seminar by Naisan Benatar: Introduction to Thermal Management

Thermal Management is an ongoing problem present in the design, maintenance and operation of modern day data centres. This presentation introduces some of the problems and some of the ways of approaching the problem, along with a brief introduction to Wireless sensor networks and how they can be used to assist in the management of the data centre. It finishes with a brief look at avenues to be explored in the domain, where they have come from and to where they might lead.

21 April 2009, 2-3pm, Room C60

Seminar by Grazziela Figueredo: Immunosenescence and its Applications to Artificial Immune Systems

The study of aging in human beings, i.e. gerontology, is a relatively new research topic and deals with understanding the processes of tissue degeneration so that they can be stopped or slowed down. Some stages of degradation indicate signs of functionality loss that precede the end of life. Even though it is not possible to prevent aging, there is a great benefit in understanding its causes, which may help to reverse some of the damage it has caused and thus improve life expectancy.

This work is concerned with the aging of the immune system (IS), known as immunosenescence. The immune system acts to fight and prevent diseases in an individual throughout life. It also interacts with other systems, i.e. the nervous, endocrine and neural systems, supporting the functioning of the whole body. However, with age there is a decay of the IS performance resulting in degenerative diseases, deregulated and ineffective immune responses. This ends up in complete collapse on the defence mechanisms resulting in death. Various theories have been proposed to explain this phenomenon, including he levels of antigenic stress, oxidation, lack of cellular resources and DNA damage.

In order to investigate how the IS ages and fails it is necessary to understand the processes by which the instabilities take place, develop, propagate and turn out to be destructive. By acquiring such knowledge it would be possible to create predictive models that can estimate the immune fitness of a patient at a certain age due to vaccination history, exposure to diseases and hemograms. Moreover, the knowledge could be applied to any other aging process, since there are many scientific, social, security, engineering, economical and computationally important problems demanding designs with more predictable degenerative-associated properties.

This seminar will present a review of the immunosenescence phenomenon including theories that could be need for modelling and developing algorithms for solving, predicting and slowing degeneration problems.

Seminar by Andrew Grundy: Routing in Wireless Networks of Varying Connectivity

Routing algorithms developed for connected environments fail in disconnected environments, as most disconnected scenarios fail to meet the assumptions of these protocols. Commonly, data packets are epidemically disseminated (controlled or restricted flooding) throughout disconnected networks. The algorithms developed for disconnected environments flood connected environments, sending more data packets than is necessary, causing congestion and as a consequence success ratios drop, buffers are inundated and eventually the network is crippled.

We exploit reoccurring patterns in day to day life, such as, travelling between home and work locations, learning about when we connect with others around us and capitalizing on consistent events. We use epidemic dissemination techniques to source route, not to transmit data.

I will present a functionality overview of the three components of our proposal, contact driven source routing, disconnection tolerant data forwarding and packet scheduling for reduced overheads and energy efficiency. I will describe our early emulations (simulations in ns-2 with real world data), including our evaluation of the results gathered.

7 April 2009, 2-3pm, Room C60

Tutorial by Julie Greensmith: Are we having fun yet?

Biosensing is an emerging field within pervasive computing and involves the use of 'biosensors' - lightweight devices which can monitor a variety of physiological attributes. Such attributes include heart rate/ECG, skin sweat responses, respiratory rates, muscle contraction and exerted G-forces. It is hypothesised that it is possible to ascertain a persons emotional state based on the correct interpretation of the biosensor trace data. However, such data is extremely noisy and subject to numerous sources of variation, making it a complex challenge to discriminate between emotional states. For example, an increase in heart rate could indicate either thrill or fear. Testing this hypothesis is one of the numerous activities performed by the "Thrill Laboratory", a collaboration between the MRL and Aerial, a design company, and recently IMA. The ultimate aim of the Thrill Laboratory is to produce adaptive user centric experiences, including rollercoasters which dynamically adapt depending upon how a monitored participant responds to the ride itself.

As part of this study, a large volume of data has been collected from over 60 participants on an extreme fairground ride, 'Oblivion' at Alton Towers, UK. In addition to the biosensor data we have also collected first person audio and video data, which can be correlated with the biosensed signal streams. We aim to use these datasets as a test case for the classification of emotions in a real world environment. My role in the project is to co ordinate the analysis of collected sensor data, model the processes involved and develop intelligent techniques to analyse this data in near to real-time. As part of this talk I will introduce the Thrill Laboratory, demonstrate the equipment used, provide an overview of the data collected to data, and detail the research we are currently performing with this data. This talk will conclude with insights into how we intend to use this technology in the near future.

31 March 2009, 2-3pm, Room C60

Seminar by Hazlina Hamdan: Modelling Survival Prediction in Medical Data

Medical prognosis is the principal factor in estimating of cure, complication, disease recurrence or survival for a patient or group of patients after treatment. Prognosis is important because the type and intensity of the medication are based on it.

Survival analysis describes the analysis of data that corresponds to the time from when an individual enters a study until the occurrence of some particular event or end-point. It is concerned with the comparison of survival curves for different combinations of risk factors. Survival data are conveniently summarised through estimates of two functions known as the survival function and the hazard function. Analysis is complicated by the presence of censorship (patients leaving the study). Statistical methods are commonly used in the analysis of survival data and lately Artificial Neural Network (ANN) have been considered as alternative methods.

In this talk, I will present the essential principles behind survival analysis, detail a previous ANN approach in modelling survival analysis and demonstrate an example from clinical data.

17 March 2009, 2-3pm, Room C60

Tutorial by Robert (Bob) Oates: An Introduction to Cybernetics

This tutorial will provide a brief introduction to the field of cybernetics, including: a definition of this much misunderstood term; the origins of the topic; its relationship to control theory and application areas that may be of interest to the group. The talk will end with an exploration of a famous cybernetic model, James Lovelock's "Daisyworld" and demonstrate the differences between a traditional simulation based on differential equations and an agent-based simulation.

10 March 2009, 2-3pm, Room C60

Tutorial by Jianyong Sun: A Gentle Introduction to Latent Variable Model and Variational EM Algorithm

Latent variable model (LVM) is widely applied as a powerful approach to probabilistic modelling. LVM involves supplementing a set of observed variables with additional latent, or hidden variables. If defining a joint distribution over the observed and latent variables, the corresponding distributions of the observed variables can be obtained by marginalization over latent variables. LVM allows relatively complex distributions to be expressed for wider application. To estimate the parameters using maximum likelihood estimation (MLE), the most widely applied methods are the expectation-maximization (EM) algorithm and its generalization, the variational EM algorithm. In this tutorial, I will first briefly introduce an overview of latent variable model for representing continuous variables. The structural of LVM can be represented by a graphical representation, usually in terms of directed acyclic graph, or Bayesian network. Secondly, the EM algorithm and variational EM algorithm are presented, and the EM algorithm is demonstrated by a well-known LVM, the Gaussian mixture model for clustering. Finally, I present a robust mixture modelling in the presence of measurement errors, including the proposed latent variable model and the variational EM for the corresponding parameter estimation.

24 February 2009, 2-3pm, Room C60

Seminar by Linda Fiaschi: Standards for SNPs Analysis with Decision Trees Tools

Data mining is still considered a not well established field as there is no standard procedure leading from data to models.

The methodology for data analysis concerns lots of interesting remarks as the choice of an adequate early sequence of steps in data exploration for modelling. It includes the selection of the features, the model and the algorithm with its parameters, the proper pre-processing, the splitting of the data, the validation test, how to treat missing, unbalanced or non standard data etc. Despite lots of methods are nowadays available it is not clear yet which one to choose for a specific application.

In this talk I will propose a sort of guideline that can help data miners in making relevant choices on the specific field of decision tree algorithms applied to the SNPs analysis for pre-eclampsia disease.

Seminar by Jan Feyereisl: Cluster Interpretation of the Self-Organising Map (Part 2/2)

The Self-Organising Map algorithm developed by Teuvo Kohonen is an established unsupervised machine learning technique that maps high-dimensional data onto a lower-dimensional representation while maintaining topographic properties of the input data within the output representation. The inherent property of organisation, out of possibly total disorder, is one of the most interesting aspects of the algorithm that captivated researchers in various fields of science. This unique property has ensured the algorithm's success, particularly within the fields of high-dimensional data visualisation and clustering.

Even with the algorithm's apparent success, the interpretation of the lower dimensional maps is still challenging and not ideal. A number of methods for better understanding of the generated maps have been developed in the past, however these are either combinations of established clustering techniques or techniques as distant as image recognition.

We propose to exploit the self-organising properties of Kohonen's algorithm in combination with a biologically inspired concept for the purpose of better clustering and subsequent interpretation of maps generated by Kohonen's original algorithm.

17 February 2009, 2-3pm, Room C60

Tutorial by Milena Radenkovic: Experiences with Supporting Mass Socio Technical Sensor Networks

There has been a proliferation of interest in ad hoc networking for providing communications among users and devices where essentially no pre-planned infrastructure is present. The large majority of the state of art research has focused on wireless sensor networks in remote areas, military and emergency response scenario, and interplanetary networking. In contrast, this talk is concerned with different ways in which many humans, animals and sensors can collaborate as part of a mass socio technical sensor network. We describe a number of projects that explore how pervasive computing can support large scale sensory applications such as mass participatory campaigns in which millions of people gather and share information about their local environments, continuous monitoring of animal wellbeing and human health. The challenges facing the networking community for supporting distributed storage and processing of information in such mass-scale applications are fundamentally profound, especially where they involve mobile sensors, diverse sensing technologies, span multiple organizations and disconnected regions. We discuss the applicability of P2P overlays to help build applications on the top of mobile and environments, and the suitability of delay tolerant network approaches to routing to help provide more reliable data communication.

10 February 2009, 2-3pm, Amenities Building, Room B12

Seminar by Jan Feyereisl: ClusterIinterpretation of the Self-Organising Map (Part 1/2)

The Self-Organising Map algorithm developed by Teuvo Kohonen is an established unsupervised machine learning technique that maps high-dimensional data onto a lower-dimensional representation while maintaining topographic properties of the input data within the output representation. The inherent property of organisation, out of possibly total disorder, is one of the most interesting aspects of the algorithm that captivated researchers in various fields of science. This unique property has ensured the algorithm's success, particularly within the fields of high-dimensional data visualisation and clustering.

Even with the algorithm's apparent success, the interpretation of the lower dimensional maps is still challenging and not ideal. A number of methods for better understanding of the generated maps have been developed in the past, however these are either combinations of established clustering techniques or techniques as distant as image recognition.

We propose to exploit the self-organising properties of Kohonen's algorithm in combination with a biologically inspired concept for the purpose of better clustering and subsequent interpretation of maps generated by Kohonen's original algorithm.

Seminar by Pooja Jain: Supervised Machine Learning Algorithms for Protein Structure Classification

In this talk, I will discuss the automation of protein structural classification using supervised machine learning methods. Wherein, fifteen algorithms from five categories of supervised algorithms are evaluated for their ability to learn for a pair of protein domains, the deepest common structural level within the SCOP hierarchy, given a one dimensional representation of the domain structures. This representation encapsulates evolutionary information in terms of sequence identity and structural information characterising the secondary structure elements and lengths of the respective domains. The evaluation is performed in two steps, first selecting the best performing base learners and subsequently evaluating boosted and bagged meta learners. The boosted random forest, a collection of decision trees, is found to be the most accurate, with a cross-validated accuracy of 97.0% and F-measures of 0.97, 0.85, 0.93 and 0.98 for classification of proteins to the Class, Fold, Super-Family and Family levels in the SCOP hierarchy. The meta learning regime, especially boosting, improved performance by more accurately classifying the instances from less populated classes.

SCOP: http://www.bio.cam.ac.uk/scop/
Random Forests: http://en.wikipedia.org/wiki/Random_forest

3 February 2009, 2-3pm, Room C60

Seminar by William Armitage: Clusters and Conferencing

This talk covers two areas. The first section covers the main university cluster, what it should provide and how it is progressing. The second section looks at some of the tools available for multi-site collaboration.

27 January 2009, 2-3pm, Room C60

Tutorial by Jon Garibaldi: Introduction to Fuzzy Logic

Fuzzy logic was introduced by Lotfi Zadeh in the 1960s and has subsequently become one of the three main pillars of 'soft computing' (a term often used to cover fuzzy logic, neural networks and evolutionary computing). It has spawned a huge area of academic research and, more recently, has made an impact in consumer electronic devices, such as 'fuzzy logic washing machines', 'fuzzy logic auto-focus cameras' and 'fuzzy logic automatic gearboxes'. This talk will present an overview of fuzzy sets, which underpin fuzzy logic, fuzzy control and fuzzy inference. More recent developments such as type-2 fuzzy logic and other non-standard forms of fuzzy sets will be briefly discussed. The talk will be aimed at novices without previous knowledge of fuzzy logic, who wish to gain a basic understanding of the field.

20 January 2009, 2-3pm, Room C60

Seminar by Qi Chen: Dempster-Shafer for Classification and Anomaly Detection

The Dempster-Shafer (D-S) theory is a mathematical theory of evidence, which becomes more popular and has been used in many research areas in the past 30 years. This work has utilized the D-S theory (in particular mass functions and Dempster's rule of combination) as a framework for creating classification systems and anomaly detection systems. This thesis focuses on the mass functions and the Dempster's rule of Combination, which are two important aspects of the D-S theory. Several mass functions are developed. Two D-S applications to real world problems have been presented. One application aims to identify the cause of energy power outage, and the other application aims to detect network anomaly.

Experiments results are compared with several traditional classification algorithms, such as logic regression, neural network, E-algorithm, and several artificial immune systems. The results presented in this thesis indicate that the D-S algorithm can solve complex real world problems; it is suitable for imbalanced data, and nominal data. The D-S algorithm can detect network anomaly in real time with high true positive rates and high true negative rates. Following the analysis, A D-S architecture and a mass function have been suggested for nominal data. The experimental results showed that the suggested mass function for nominal data can be optimized, which indicate that the function may be optimized to fit for other nominal datasets.

16 December 2008, 9.30am – 3.00pm, Room B52

Short Course by Will Lowe: Statistics for Computer Scientists

Statistics for computer scientists is a short course in the principles of statistical analysis and experimentation or applications in computer science. It is designed as a broad and introductory overview to help students:

1. understand the logic of statistical and experimental inferences,
2. interpret statistical results presented in computer science papers,
3. determine appropriate methods to analyse their own data, and
4. communicate the results of their research clearly and effectively.

The course is primarily focused on providing students a clear theoretical understanding of the material but here are also small practical exercises in the R language demonstrating various methods of analysis. A detailed breakdown of the topics covered can be found on the following page. There is no course textbook. Recommendations for further reading will be made at the end of the course.

9 December 2008, 2-3pm, Room C60

Tutorial by Jon Garibaldi: An Introduction to Statistics for Computer Scientists

The use of formal statistical procedures in experimental work is becoming the accepted standard in most areas of computer science. This requires the formal statement of hypotheses, basic understanding of data types and distributions, and statistical hypothesis testing. In this talk, I will present the essential concepts required in order to gain an elementary understanding of this field and provide pointers as to how and where to find more information for specific requirements. Topics will include types of data, distributions, parametric and non- parametric methods, null and alternate hypotheses and p-values. Concepts will be illustrated will practical demonstrations.

2 December 2008, 2-3pm, Room C60

Seminar by Amanda Whitbrook: An Idiotypic Immune Network as a Short-Term Learning Architecture for Mobile Robots

A combined Short-Term Learning (STL) and Long-Term Learning (LTL) approach to solving mobile robot navigation problems is presented and tested in both real and simulated environments. The LTL consists of rapid simulations that use a Genetic Algorithm to derive diverse sets of behaviours. These sets are then transferred to an idiotypic Artificial Immune System (AIS), which forms the STL phase, and the system is said to be seeded. The combined idiotypic LTL-STL approach is compared with using STL only, and with combined and single STL systems where the idiotypic mechanism is turned off. The results provide substantial evidence that the best option is the seeded idiotypic system, i.e. the architecture that merges LTL with an idiotypic AIS for the STL. They also show that structurally different environments can be used for the two phases without compromising transferability.

25 November 2008, 2-3pm, Room C60

Seminar by Feng Gu: The Evolution of the DCA

As an immune inspired algorithm, the Dendritic Cell Algorithm (DCA) has been applied to various problem domains. Among the applications, the most promising one is intrusion detection. In practice, intrusion detection is a real-time critic mission, thus, the new challenge for the DCA is to cope with such a situation.

The newest implementation of the algorithm, DCA, features data processing in real-time, but it lacks a real-time analysis component, which can perform periodical analysis during the detection. The first step of this is to decide when to analyse, which can be accomplished by segmentation. Two segmentation approaches have been investigated, which are 'antigen based segmentation' and 'time based segmentation'. Experimental results have shown that segmentation can make significant differences to the algorithm, and antigen based segmentation performs better than time based segmentation.

Seminar by Yousof Al-Hammadi: Detecting Botnet/Bot based on Correlating Activities

For the past few years, IRC bots, malicious programs which are remotely controlled by the attacker, have become a major threat to the Internet and users by issuing distributed denial of services attack to shutdown other networks and services. Current bots have extended features such as keystroke logging, spamming, traffic sniffing cause serious disruption on networks and users. In response to these threats, there is a growing demand for effective techniques to detect the presence of bots/botnets and to reduce their malicious effect. Furthermore, existing approaches detect botnet rather than an individual bot. In this presentation, we present a behavioural approach for detecting bots/botnet based on correlating different bot's activities within a specified time window. We then extend our work to detect peer to peer bots which are the upcoming threat. Our evaluation shows that the correlating different activities generated by the bot within a specified time period not only achieved high detection accuracy, but also reduces the amount of processing network traffic compared to other signature-based detection techniques.

18 November 2008, 2-3pm, Room C60

Tutorial by Peer-Olaf Siebers: Everything you always wanted to know about Discrete Event Simulation … but were afraid to ask ;)

This will be a hands-on tutorial on (object oriented) discrete event simulation. First I will clarify some terminology used in the context of systems simulation and describe what activities are involved in a systems simulation study. After this, I will focus on discrete event modelling and simulation. You will get some hands-on experience on conducting a simulation study. First we will develop a conceptual model using active cycle diagrams. Then we will look into the internal processes during the execution of a three-phase simulation. We will also have a look at the implementation of the simulation model in AnyLogic (a multi-paradigm simulation environment). Finally, if we have the time, we will do some output analysis with the data we obtained from our simulation experiment. I hope that after this tutorial you will have an idea of what discrete event modelling and simulation is and how it works.

11 November 2008, 2-3pm, Room C60

Seminar by Mazlina Abdul Majid: Modelling Human Reactive Behaviour in Discrete Event and Agent Based Simulation

Discrete Event Simulation (DES) is a very popular simulation technique in Operational Research and Agent Based Simulation (ABS) is a new paradigm in simulation modelling. Modelling both DES and ABS in human behaviour will be addressed in this paper. We are interested in investigating these simulation techniques due to lack of literature debating about their comparison performance in human behaviour issues. In order to understand the distinctions between these two simulation techniques, our aim is to understand their output performance in relation to the real world phenomenon of modelling and simulating human behaviour. In achieving this aim, we have carried out a case study at a department store. Both DES and ABS models will be compared using the same problem domain which is concerned with management policy in a fitting room. The reactive behaviour of staff while working will be modelled using both simulation techniques to gain insight into their behaviour.

IMA Seminar on 4 November 2008, 2-3pm, Room C60

Seminar by Adrian Adewunmi: Investigating the Joint Application of Common Random Numbers and Antithetic Variates

This presentation discusses the 'stand alone' and combined application of two variance reduction techniques that can improve the reliability and efficiency of a simulation experimental process by manipulating random number seeds for each source of model variation at each replication of the simulation.

The two variance reduction techniques, Common random numbers and Antithetic variates, reduce the variance of the selected output performance measure by replacing the original sampling procedure with a new procedure that yields the same expected value but with a smaller variance. The strategy used to combine the two variance reduction techniques is based on previous work carried out by Kleijnen (1975), (1974): Common random numbers only, Antithetic variates only, and Common random numbers and Antithetic variates combined.

The application of our variance reduction techniques is illustrated using results from the simulation of a Crossdocking distribution centre. From our results, the combined application and Antithetic variates perform marginally as well as each other, followed by Common random numbers in reducing the variance of our simulation output performance measure value. These results further reinforce the opinion of Kleijnen (1975), (1974), that Antithetic variates can be chosen, a priori, as a variance reduction technique for simulation, if simulation computational effort is an initial consideration.

References:

- Kleijnen, J.P.C. (1974). Statistical Techniques in Simulation, Volume I. Marcel Dekker, Inc., New York.
- Kleijnen, J.P.C. (1975) Antithetic Variates, Common random numbers and optimal computer time allocation in simulation. Management Science, Vol. 21, No. 10, pp: 1176-1185.

28 October 2008, 2-3pm, Room C60

Tutorial by Julie Greensmith and Uwe Aickelin: How to Survive Inter-disciplinary Collaborations

We will talk about our own experiences with inter-disciplinary collaborations, what are the benefits of collaboration, the pitfalls and how to overcome them, even mention some tips on how to actually form collaborations and keep them going!

21 October 2008, 2-3pm, Room C60

Seminar by Gianni Tedesco: Real-Time Alert Correlation with Type Graphs

The premise of automated alert correlation is to accept that false alerts from a low level intrusion detection system are inevitable and use attack models to explain the output in an understandable way. Several algorithms exist for this purpose which use attack graphs to model the ways in which attacks can be combined. These algorithms can be classified in to two broad categories namely scenario-graph approaches, which create an attack model starting from a vulnerability assessment and type-graph approaches which rely on an abstract model of the relations between attack types. Some research in to improving the efficiency of type-graph correlation has been carried out but this research has ignored the hypothesising of missing alerts. Our work is to present a novel type-graph algorithm which unifies correlation and hypothesising in to a single operation. Our experimental results indicate that the approach is extremely efficient in the face of intensive alerts and produces compact output graphs comparable to other techniques.

14 October 2008, 2-3pm, Room C60

Tutorial by Robert (Bob) Oates: Introduction to Control Theory

The ability to model and analyse systems is of great interest to engineers and scientists. Even before the advent of computing, the ability to form coherent representations of abstract or non-linear systems was of interest to the scientific community and a branch of mathematics was developed to achieve just that. In this tutorial I will introduce some of the basic concepts within "Control Theory", including techniques for representing systems and predicting their outputs to arbitrary input signals. I will go on to explain the techniques required to view systems within the frequency domain and, using the dendritic cell algorithm has an example, I will demonstrate how we can form models that provide insight into the operation of algorithms.

7 October 2008, 2-3pm, Room C60

Seminar by Phil Birkin: Comparison of Tuned Membership Functions of Type-1 and Type-2 for a Seven Term Fuzzy Logic Controller

In my talk I will describe an experiment to investigate eight shapes of membership functions for a seven term Fuzzy Logic Controller to control a DC motor simulator in a closed loop configuration. Four shapes of type-1 membership functions are generated as Gaussian, Trapezoidal, Triangular and Trapezoidal-Triangular. The parameters of the membership functions are determined using a trial and error approach. Four further shapes are generated as type-2 membership functions. Each of these type-2 membership functions encloses the corresponding type-1 membership functions. All experiments/comparisons are performed with noise and without noise. Observational analysis shows that the exact shape of the membership functions used in the controllers doesn't appear to matter provided the initial controller is stable. The type-1 controllers out perform the type-2 seven term fuzzy logic controllers when no noise is applied. With noise then the type-2 controllers are better.