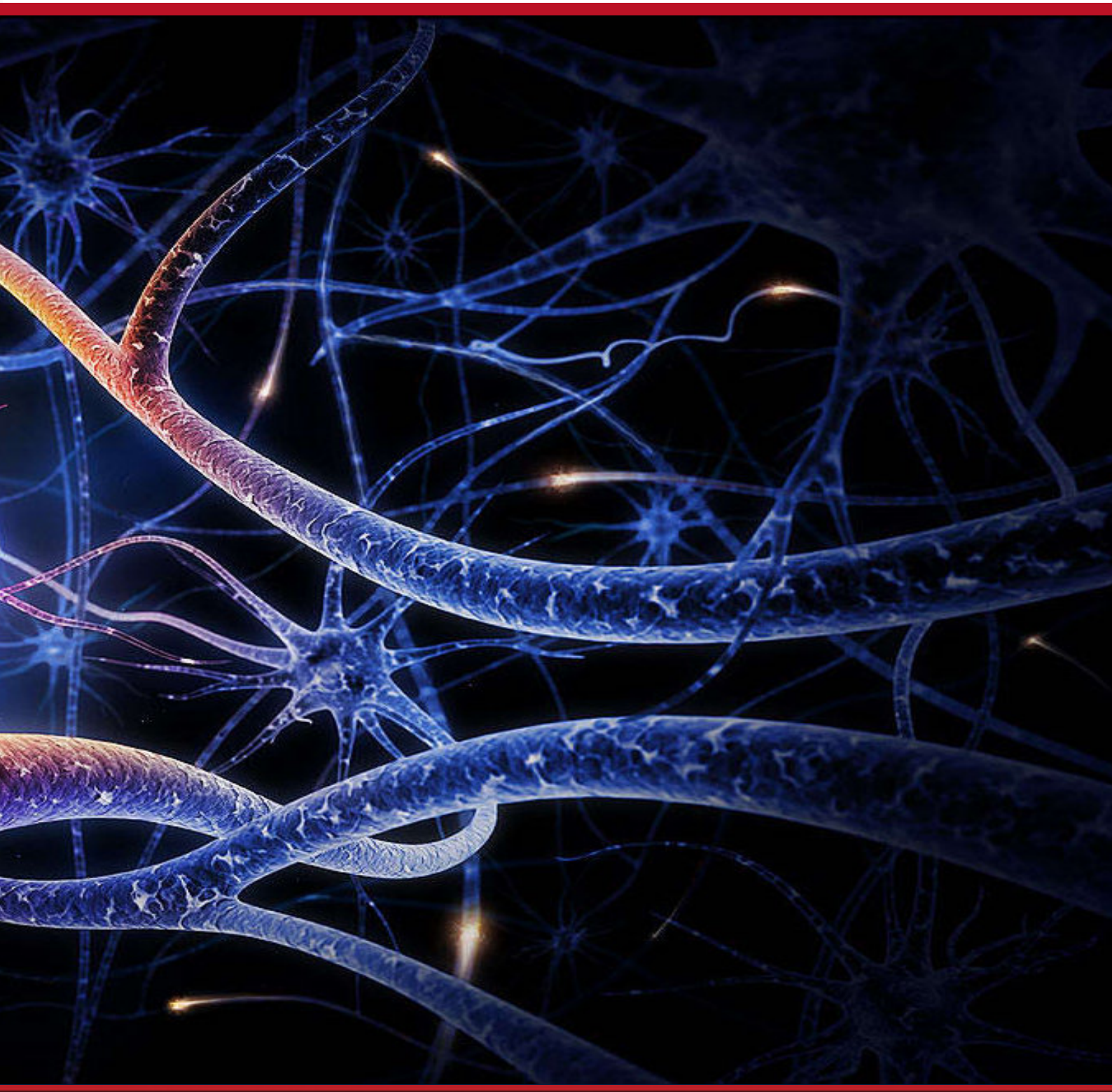
Newsletter

Operations Research Society of South Africa
Operasionele Navorsingsvereniging van Suid-Afrika



March 2011

www.orssa.org.za

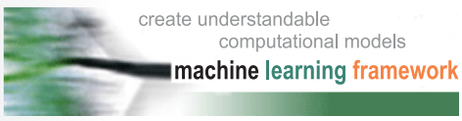
Mathematica

Mathematica v8 has more than 500 new functions in 7 new application areas, including: statistics, GPU support, wavelets, integrated image analysis, and financial applications

Wolfram Mathematica 8

Machine Learning Framework

This high performance data mining tool combines fuzzy logic, neural networks, clustering analysis, and much more.



Expert Choice

Make better, faster, more justifiable decisions with Expert Choice. Application areas include Strategic Planning, Vendor Selection, HR, Risk, and Resource Allocation

expertchoice®
otherwise you're just guessing

Systat

New and improved: More Statistics, More Graphs, Less Effort! SYSTAT has a 20 year track record of bringing you industrial strength statistics at an affordable price.

SYSTAT

LINDO

Speed and ease-of-use have made LINDO Systems a leading developer of software tools for building and solving optimization models

LINDO

BLUESTALLION technologies

www.bluestallion.co.za

TECHNICAL, SCIENTIFIC AND BUSINESS SOFTWARE

We are the proud distributors of the following products in South Africa:

SigmaPlot

The technical graphing standard used by over 100,000 researchers worldwide. Now also includes the functionality of SigmaStat

SIGMAPLOT
Exact Graphs for Exact Science

Origin

Origin is the complete solution for data analysis and technical graphing. Import data from Excel and then perform advanced analysis (including non-linear fits, FFT, ANOVA, etc)

ORIGIN

Hugin Expert

Hugin is a tool which enables you to construct decision support systems using bayesian Networks and their extension influence diagrams.

HUGIN EXPERT

MathType

MathType is a powerful interactive tool for Windows and Apple OS that lets you create mathematical notation for word processing, web pages and presentations

MathType

AnyLogic

Anylogic is a professional simulation tool for complex discrete, continuous, hybrid and agent based systems. Powerful and flexible, it is used to model, simulate, visualize and analyze a diverse range of problems

AnyLogic® 6

For a free evaluation CD and brochures, please call (011) 447 9916 or email: info@bluestallion.co.za
PO Box 786246, Sandton, 2146, South Africa Tel: (011) 447 9916 Fax: 011 447 9911

FROM THE EDITOR

Contactable at: 14556561@sun.ac.za

Danie Lötter

Greetings to all the ORSSA members. I hope that you all had a wonderful holiday. This edition starts off with the usual column from the President's Desk. A summary of the new executive committee for 2011 may also be found on the same page. A student competition was hosted by the Vaal Triangle chapter (a first for this chapter) and a short article on this follows. The featuring article in this edition is titled *Spam detection using generalized neural networks* by Tiny du Toit and André de Waal from the North-West University.

The member interview for this edition is conducted with Winny Pelser from Armscor's Defense Decision Support Institute. By the way, she is also the chair of the Pretoria chapter. A new online submission system for ORiON has been tested live in 2010. This system is now fully operational and guidelines for the submission process may be found on page 13.

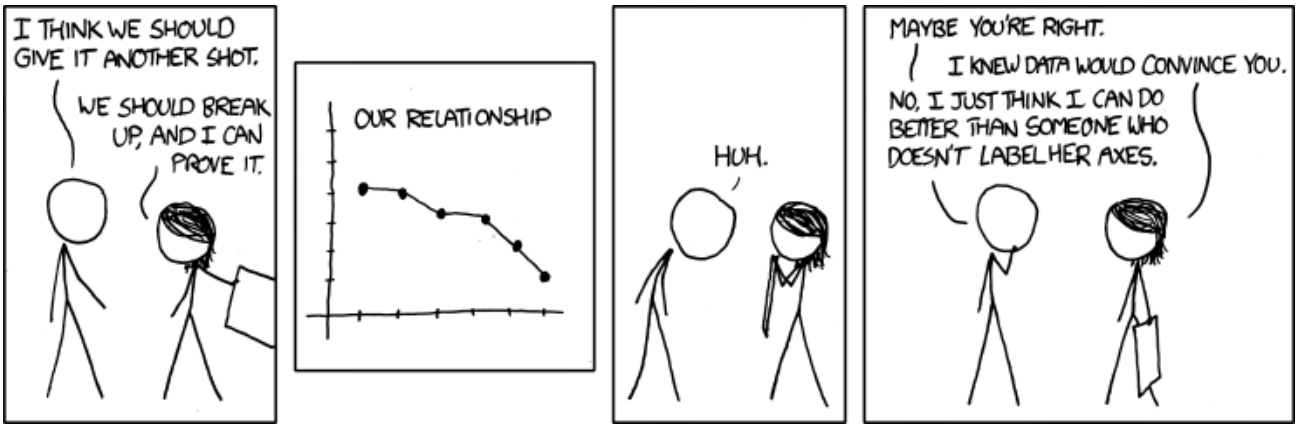
Then I would like to urge members to read the notices on pages 9, 15 and 16. These are important announcements regarding submissions for the student competition, the Tom Rozwadowski award and information regarding the annual conference. Enjoy the first issue of 2011☺.

Features

QUERIES AND CONTRIBUTIONS

Any queries and contributions to the newsletter are most welcome, especially article submissions. For any queries and contributions, please contact the newsletter editor: Danie Lötter

Email: 14556561@sun.ac.za



FROM THE PRESIDENT'S DESK

by Dave Evans (davee@dbsa.org)
ORSSA President

A big welcome to ORSSA members to the new year. Let's make it a special one for the Society. We've already taken one big step in that direction by deciding to hold this year's conference, our fortieth, in Zimbabwe, at the Elephant Hills Hotel, at Victoria Falls.

The Executive Committee has been discussing the idea of an 'African Outreach' for a while, and everyone who was at last year's conference in Limpopo will be aware that the idea of having the 2011 conference in Zimbabwe was very well received, and encouraged by our friend and keynote speaker, Prof Jim Cochran.

In holding the conference in Zimbabwe, we hope to foster greater collaboration and cooperation across southern African OR communities, accelerate the expansion of OR applications and education in Zimbabwe, and encourage the establishment of a Zimbabwean OR society. Prof Cochran is publicising the conference via the American OR journal channels he is involved in, and Elise del Rosario, past president of IFORS (and now their newsletter editor) is doing the same, so we may have a greater than usual 'non-African' participation, in addition to the wider spread of African delegates we hope to attract.

With the theme of "Spreading Operations Research Across Africa," we are moving in our stated direction and working to expand the awareness, understanding, and use of OR across the wider region. This theme encourages participation over the full spectrum of Operations Research, welcoming papers of a more fundamental nature, those on the application of Operations Research techniques in business and industry, and those about topical issues in Operations Research, and about the philosophy, teaching and marketing of Operations Research.

The Conference will be hosted by the National University of Science and Technology, Bulawayo, Zimbabwe. A Local Organising Committee is already in full swing, supported from here by Prof. 'Maseka Lesaoana of the University of Limpopo and Ozias Ncube from UNISA, both of whom were heavily involved in last year's conference. It will take place from 18th to 21st September 2011, with 22nd September reserved for sightseeing.

I'd also like to focus this month on the outstanding job which Jan van Vuuren and his team have done on our 'publications' over the past few years.

NEW ORSSA EXECUTIVE COMMITTEE 2011

President:
Dave Evans

Vice-President:
Jan van Vuuren

Secretary:
Isabelle Nieuwoudt

Treasurer:
Marthi Harmse

ORiON editor:
Stephan Visagie

ORiON Business Manager:
Martin Kidd

Newletter editor:
Danie Lötter

Newsletter Business Manager:
Francois Bester

Webmaster:
Basie Kok

Additional members:
Ozias Ncube 'Maseka Lesaoana
Elias Willemse Ian Durbach

Chapter Chairs:
Johannesburg: Louis Dannhauser
Pretoria: Winnie Pelser
Vaal Triangle: Hennie Kruger
Western Cape: Margarete Bester

IFORS Representative: Hans Ittmann

Jan took over as the editor of ORiON in 2004, and for seven years, he has steadily taken both it and the Newsletter to progressively higher levels. This in no way takes anything from the previous journal editors, Theo Stewart, the founder editor in the five years from 1985 to 1989, Marius Sinclair from 1990 to 1993, Yvonne Walus during the period 1994-5, and Paul Fatti from 1996 to 2003.

Jan assembled a strong team in Stellenbosch, including Stephan Visagie as Business Manager, Lieschen Venter, Anton de Villiers, Adri van der Merwe and Martin Kidd who have done a sterling job on our journal for the past few years. Unless you've been in that game yourself, it is impossible to appreciate the amount of work involved in publishing our newsletter, let alone the journal. Throughout their tenure, Jan and the team have professionally and pleasantly encouraged contributors and referees, and managed publishers in a totally effective manner, the results being the high quality publications which we regularly receive and enjoy. Similarly, Danie Lötter and Francois Bester have been doing an outstanding job on the Newsletter.

Jan did not 'resign' from the editorship willingly; the Executive Committee felt that he was a natural choice to be our next President, when my term expires at the end of 2011, but his commitment to ORiON and his humility meant that I had to cajole him remorselessly to get him to accept his current post of vice-president.

One of his conditions for relinquishing his day to day editorship of ORiON was that he had strong views on who should succeed him. Fortunately (and predictably, given the quality and experience of the people involved) the Executive Committee had exactly the same views – the changes in the team selected themselves, and after the relevant discussions with the people affected, a very logical and smooth transition has been affected. Congratulations to Stephan on being elected to the post of Editor, and to the rest of his team, who I am sure will continue with the excellent job they have been doing on the journal and newsletter.

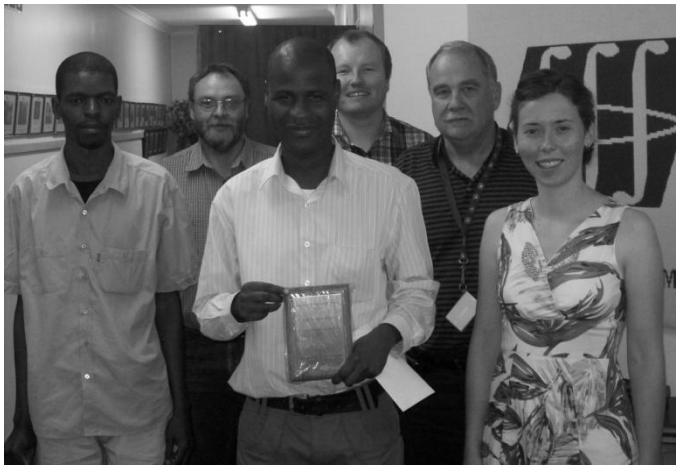
I can't overestimate the vote of thanks we owe to Jan and his team. And whilst he is now vice-president, I'm sure there will still be interactions between him and the rest of the team. I look forward to working with them all in the coming months, as we have done in the past, and to Jan's period as president in 2012 and 2013.

VAAL TRIANGLE CHAPTER: STUDENT COMPETITION

by Hennie Kruger (hennie.kruger@nwu.ac.za)

A successful student competition was held on 26 November 2010 by the Vaal Triangle chapter of ORSSA. Students who completed their masters or PhD studies during 2010 on an OR or OR-related topic were invited to participate. Three entries were received. The students competed by giving a 20-30 minute presentation of their work which was judged by two independent judges. The event and the prizes were sponsored by the Centre for Business Mathematics and Informatics at the North-West University. The two judges were Prof Riaan de Jongh (Centre for BMI) and Dr James Allison (School for Computer, Statistics and Mathematical Sciences). The winner was Tumo Baitshenyetsi who is a Masters student at the North-West University (Potchefstroom Campus). His presentation was titled, *Applying tree knapsack approaches to general network design: a case study*. Tumo received a shield from ORSSA as well as a cash prize of R2500.

The Vaal Triangle Chapter would like to express their gratitude towards the sponsor, the judges and the participants. The competition was a first for the Vaal Triangle chapter and the intention is to make this a regular annual event on the Vaal Triangle ORSSA calendar.



From left to right is, Sanele Manyatsi (participant), Hennie Kruger, Tumo Baitshenyetsi (winner), James Allison, Riaan de Jong and Magderie van der Westhuizen (participant).

Spam detection using generalized additive Neural Networks

by Tiny du Toit[†] (tiny.dutoit@nwu.ac.za) and André de Waal (andre.dewaal@sas.com)*
†School of Computer, Statistical and Mathematical Sciences, North-West University
***Centre for Business Mathematics and Informatics, North-West University**

Since the late 1990s the quantity of email sent has grown exponentially. Moreover, the amount of spam (the Concise Oxford English Dictionary defines spam as "irrelevant or inappropriate messages sent on the Internet to a large number of newsgroups or users.") has increased even more. In 1998 approximately 10% of the overall mail volume was comprised of spam. By 2007 this number has increased to as much as 80% (Cranor and LaMacchia, 1998); (Goodman, Cormack and Heckerman, 2007). More than a billion spam messages are sent daily to large email services such as Microsoft's Hotmail. This deluge of unsolicited messages creates a heavy burden on both tens of millions of end users worldwide and tens of thousands of email service providers (ESPs). Spam takes away resources from users and service providers without providing any remuneration or obtaining authorization (Kiran and Atmosukarto, n.d.). Spam emails are normally sent using bulk mailers and address lists that are acquired from web pages and newsgroup archives. Their content range from deals to real estate to pornographic material.

In this article the proliferation of spam is attacked by a relatively new type of neural network. Generalized additive neural networks have a number of favourable properties which provide grounds for an investigation into the domain of spam detection. An automated construction algorithm has been developed which utilizes a greedy best-first search procedure that identifies good models in short time periods. These models proved to have high predictive accuracy and are comparable to other models found in the literature that distinguish between spam and good emails. With the automated algorithm, in-sample model selection, cross-validation, and feature selection can be performed.

Nearly all spam filtering systems utilize at least one machine learning component (Goodman et al., 2007), where computer programs are presented examples of both spam and good email. The characteristics of the spam email versus the good email are then determined by a learning algorithm. Accordingly, future incoming messages can be automatically classified as probably spam, probably good, or somewhere in between.

Learning approaches were initially fairly simple and used techniques like the Naive Bayes algorithm to count how frequently each feature or word appeared in spam messages or good messages. Naive Bayes and other similar techniques require training data - known spam and known good mail – to train the system. When spam was becoming a major problem around 1998, it was relatively static. A trained filter did not need to be updated for a number of months. Certain words like "free" or "money" were sufficient indicators of spam and functioned for a lengthy period. Unfortunately, spammers adapted to the more widely deployed spam filters. They quickly learned the most obvious words to avoid and the most innocent words to add to lead the filter astray. To keep up with the spammers, it became necessary to collect increasing amounts of email as spammers made use of a wider variety of terms. Filters also had to be updated frequently. Currently, Hotmail uses more than 100,000 volunteers who are daily asked to label an email that was sent to them as either "spam" or "good" email. This feedback loop system provides Hotmail with new messages to train their filters, allowing them to respond quickly to new spammer attacks and schemes.

Apart from getting more training data, faster, much more advanced learning algorithms are currently being used. For example, algorithms based on logistic regression and support vector machines can bring down the amount of spam that bypass filtering by half, compared to Naive Bayes (Goodman et al., 2007). With these algorithms the messages are broken down into individual words and weights are "learned" for each word. The weights are carefully adjusted to obtain the

most accurate results from the training data. The learning process can be a potentially time-consuming operation as tens of thousands or even hundreds of thousands of weights may require repeated adjusting. Fortunately, such computation has been made possible by advances in machine learning over the past few years. Complex algorithms like Sequential Conditional Generalized Iterative Scaling allows Hotmail to learn a new filter from scratch in about one hour with training data of more than a million emails.

In the following section a recently developed neural network architecture is employed to detect spam. This design does not suffer from the black box perception ascribed to artificial neural networks in general as visual diagnostics provide insight into the models created. Furthermore, no user input is required while an automated algorithm searches for the best model.

GENERALIZED ADDITIVE NEURAL NETWORKS

Spam detection can be regarded as an instance of the generic supervised prediction problem which consists of a data set having a number of cases (messages) (Potts, 1999). Each case is associated with a vector of input variables (features) $x_1, x_2, \cdots, x_k$ and a target variable *y*. The latter represents a class label that indicates whether an email is spam or non-spam. A predictive model maps the inputs to the expected value of the target and is built on a training set where the target is known. The objective is to apply the model to new data where the target is unknown. Generalized linear models (McCullagh and Nelder, 1989) of the form,

$$g_0^{-1}(E(y)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k,$$

are often used for predictive modeling. The range of predicted values are restricted by the link function, g_0^{-1}. For spam detection, the logit link

$$g_0^{-1}(E(y)) = \ln(\frac{E(y)}{1-E(y)}),$$

is appropriate as the expected target (probabilities) is bounded between zero and one. The parameters are usually estimated by maximum likelihood.

Multilayer perceptrons (Bishop, 1995); (Ripley, 1996); (Zhang, Patuwo and Hu, 1998) are the most widely used type of neural network for supervised prediction. A *multilayer perceptron* (MLP) with a single hidden layer with *h* hidden neurons has the form

$$g_0^{-1}(E(y)) = w_0 + w_1 \tanh(w_{01} + \sum_{j=1}^{k} w_{j1} x_j) + \cdots + w_h \tanh(w_{0h} + \sum_{j=1}^{k} w_{jh} x_j),$$

where the link function is the inverse of the output activation function. Although other sigmoidal functions may be used, the activation function in this case is the hyperbolic tangent. The unknown parameters are estimated by numerically optimizing some appropriate measure of fit to the training data such as the negative log likelihood.

A *generalized additive model* (GAM) is defined as $g_0^{-1}(E(y)) = \beta_0 + f_1(x_1) + \cdots + f_k(x_k)$ where the expected target on the link scale is expressed as the sum of unspecified univariate functions (Hastie and Tibshirani, 1986); (Hastie and Tibshirani, 1990); (Wood, 2006). Each univariate function can be regarded as the effect of the corresponding input while holding the other inputs constant. When a GAM is implemented as a neural network it is called a *generalized additive neural network* (GANN).

The main architecture of a GANN is comprised of a separate MLP with a single hidden layer of *h* units for each input variable,

$$f_j(x_j) = w_{1j} \tanh(w_{01j} + w_{11j} x_j) + \cdots + w_{hj} \tanh(w_{0hj} + w_{1hj} x_j).$$

The individual bias terms of the outputs are incorporated into the overall bias β_0. Each individual univariate function contains 3*h* parameters, where *h* may be different across inputs. This architecture can be extended to include an additional parameter for a direct connection (skip layer),

$$f_j(x_j) = w_{0j} x_j + w_{1j} \tanh(w_{01j} + w_{11j} x_j) + \cdots + w_{hj} \tanh(w_{0hj} + w_{1hj} x_j).$$

A backfitting algorithm is used by Hastie and Tibshirani (1986); Hastie and Tibshirani (1990) to estimate the individual univariate functions f_j. Backfitting is not required for GANNs. Any method that is suitable for fitting more general MLPs can be utilized to simultaneously estimate the parameters of GANN

models. The usual optimization and model complexity issues also apply to GANN models.

Presently two algorithms exist to estimate GANN models. Potts (1999) suggested an interactive construction algorithm that makes use of visual diagnostics to determine the complexity of each univariate function. Plots of the fitted univariate functions, $\hat{f}_j(x_j)$, overlaid on the partial residuals

$$pr_j = g_0^{-1}(y) - \beta_0 - \sum_{l \neq j} \hat{f}_l(x_j) = (g_0^{-1}(y) - g_0^{-1}(\hat{y})) + \hat{f}_j(x_j)$$

versus the corresponding *j*th input are utilized for model selection (Berk and Booth, 1995); (Ezekiel, 1924); (Larsen and McCleary, 1972). When GANNs are constructed interactively, human judgment is required to interpret the partial residual plots. For a large number of inputs this can become a daunting and time consuming task. Also, human judgment is subjective which may result in the creation of models that are suboptimal. Consequently, Du Toit (2006) developed an automated method based on the search for models using objective model selection criteria or cross-validation. With this approach, partial residual plots are not used primarily for model building, but as a tool to provide insight into the models constructed. When given adequate time to evaluate candidate models, this best-first search technique is complete and optimal. Du Toit showed that the algorithm is powerful, effective and produces results comparable to other non-linear model selection techniques found in the literature.

In the next section the implementation of the automated construction algorithm, called *AutoGANN*, is used to classify incoming email into spam or good messages.

EXAMPLE

The Spambase data set (Asuncion and Newman, 2007) has 4,601 instances where each instance denotes a single message and 39.4% are classified as spam. There are 57 continuous non-missing inputs and a binary target indicating spam (1) or non-spam (0). Most of the inputs (54) indicate how frequently a particular word or character occurred in each email and was encoded as a percentage in [0, 100]. Examples of words and characters are “business”, “credit”, “edu”, “free”, “internet”, “!”, “#” and “$”. Finally, there are three run-length inputs that measure the length of sequences of consecutive capital letters.

Kiran and Atmosukarto (n.d.) performed a number of experiments on the Spambase data set to analyze various implementation and design aspects of spam filtering. They considered eight classification algorithms, namely decision trees, support vector machines, Naive Bayes, neural networks, ensemble decision trees, boosting, bagging and stacking. For all the experiments the data were partioned into a training set and a testing set. The latter was unseen by the classifiers and performance was measured by evaluating the accuracy (Table 1). Although no indication was given with respect to how the data were partitioned, additional experiments were conducted to determine whether the accuracies could be improved. A random 50% - 50% split into spam and good emails for the training set, a leave-one-out cross-validation and *k*-fold cross-validation with *k* = 10 were carried out without significant improvements from the above results. It was decided to split the data into 70% (training) and 30% (testing) subsets. The AutoGANN system constructed a model with an accuracy of 94.28%, thereby establishing a third place in the list of classifiers in Table 1. Of the 42 inputs selected, 31 were identified as having linear relationships with the target and 11 inputs had non-linear relationships with the target. Examples of inputs (words) removed from the model are *addresses*, *direct*, *mail*, *people*, *table* and *your*. It would seem as if these words do not differentiate between spam and good emails.

Classifier	Accuracy(%)
Ensemble decision tree	96.40
Adaboost	95.00
Stacking	93.80
Support vector machine	93.40
Bagging	92.80
Decision tree	92.58
Neural network	90.80
Naïve Bayes	89.57

Table 1: Classifier accuracy results.

Although the Ensemble decision tree and the Adaboost methods produce more accurate results than the AutoGANN system, their results are very difficult to interpret. The Ensemble decision tree method creates a magnitude of small models and combines the results into one complicated final model. The separate models, from which the final model is constructed, are usually not available and the results are therefore very difficult to interpret. Adaboost calls a classifier (such as a decision tree method) repeatedly, where incorrect classified cases are given more weight in subsequent iterations/models. The final model is also difficult to interpret as the individual models constructed during each iteration of the algorithm are not available and the resulting final model may by very complex. The AutoGANN system loses some predictive accuracy over that of the Ensemble decision tree and Adaboost methods, but it is a price worth paying for the increased interpretability that is further elaborated on in the next section.

DISCUSSION

The automated construction algorithm solves the problem of architecture selection by organising the GANN models into a search tree and performing a greedy best-first search (De Waal and Du Toit, 2007). Out-of-sample performance or an in-sample model selection criterion can be optimized. In addition, two heuristics are applied to speed up the search. First, a stepwise regression identifies significant inputs and their relationships with the target. This information is combined to create a clever starting point (GANN model) from which the search can commence. Basically, this heuristic performs an intelligent guess of the best architecture. The better the guess, the less searching must be performed to obtain the best model. In the example, this particular GANN model achieved an accuracy of 91.03%, outperforming the Naive Bayes and neural network classifiers, as shown in Table 1. One-half (21) of the 42 inputs selected in the best GANN model, were already identified by the intelligent start. A second heuristic allows multiple changes to successive architectures examined. This rule of thumb enables the algorithm to make systematic leaps in the search tree.

Neural networks are usually regarded as black boxes with respect to interpretation. The influence of a particular input on the target can depend in complicated ways on the values of the other inputs. In some applications, such as voice recognition, pure prediction is the goal; understanding how the inputs affect the prediction is not important. In many scientific applications, the opposite is true. To understand is the goal, and predictive power only validates the interpretive power of the model. Some domains, such as spam detection, often have both goals. Evaluating new cases is the main purpose of predictive modelling. However, some understanding, even informal, of the factors influencing the prediction can be helpful in making progress towards developing better spam filters.

Figures 1, 2, 3 and 4 present partial residual plots for the inputs *edu*, *free*, *hp* and *internet* respectively. These diagrams allow the modeller to gain insight into the constructed model. All four inputs were identified as having non-linear relationships with the target (non-straight lines in the partial residual plots). Figure 1 shows a reverse trend between the frequency of the word *edu* and the probability that the email is spam. On the other hand, Figures 2 and 4 indicate that an increase in the frequencies of the words *free* and *internet* raises the probability of the email being classified as spam. Figure 3 shows a sharp reduction in probability for small frequencies of the word *hp*, followed by a constant probability for increasing frequencies. Although a non-linear relationship between *internet* and the target was identified, Figure 4 emphasises that a linear relationship would also suffice. Such a modification would result in a more parsimonious model. In the next section, some final conclusions are made.

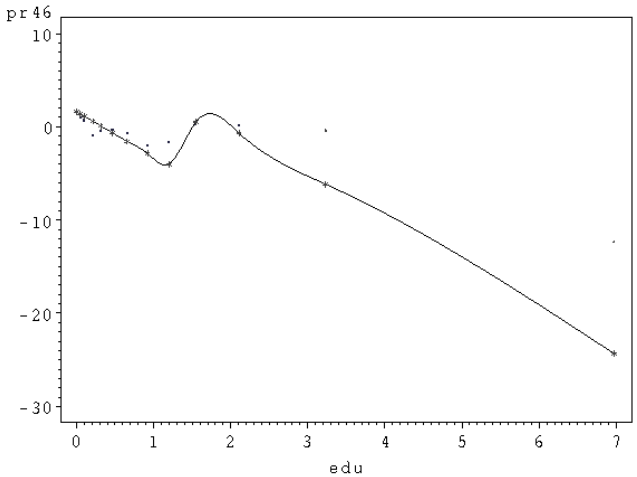


Figure 1: Partial residual plot for *edu.*

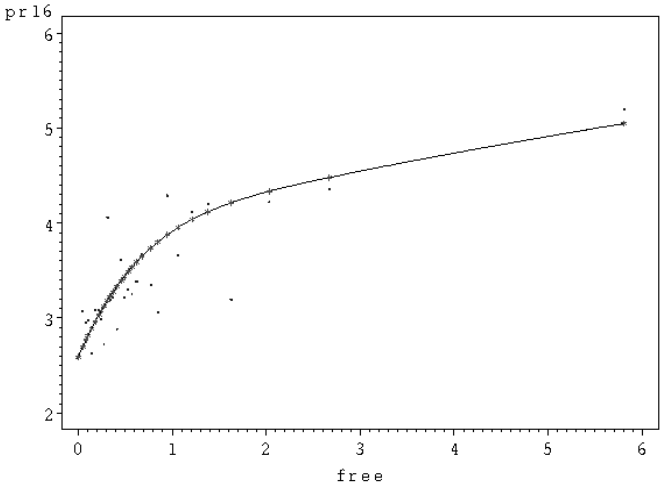


Figure 2: Partial residual plot for *free.*

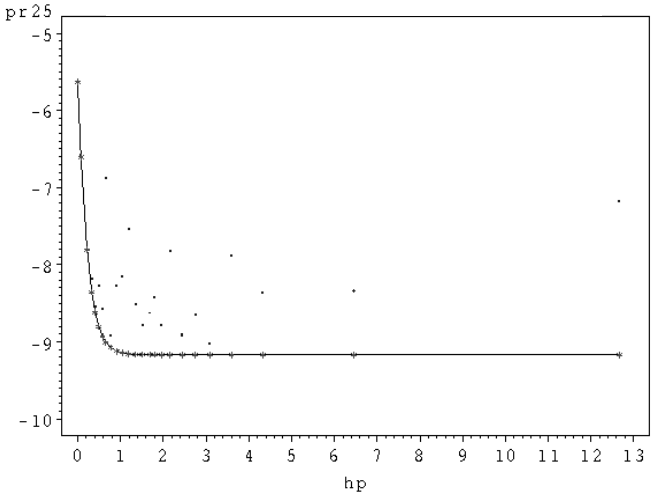


Figure 3: Partial residual plot for *hp.*

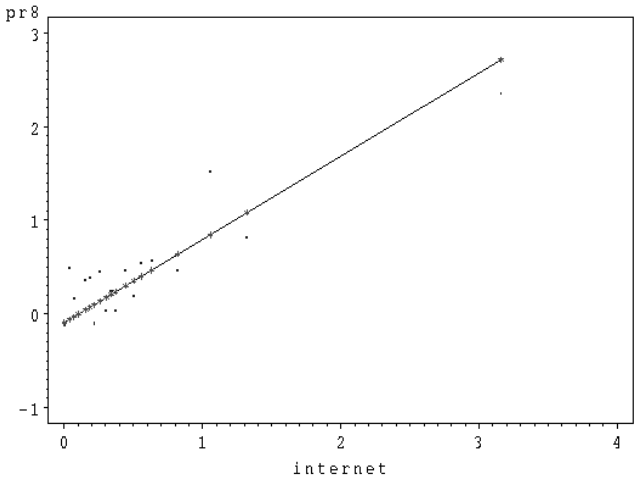


Figure 4: Partial residual plot for *internet.*

CONCLUSIONS

Scientific evaluation is a crucial component of research as researchers must be able to compare methods using standard data and measurements. This type of evaluation is particularly difficult for spam filtering. Building a standard benchmark for use by researchers is difficult in view of the sensitivity of email. Few organisations and individuals would allow their own messages to be publicly shared and those that would are hardly representative. This predicament is exemplified by the Spambase data set. Although this particular data set serves as an adequate testbed for evaluating a new spam filter, the inputs were selected on the basis of email arriving at one specific individual at one specific corporate organization. As a consequence the attributes are not representative of a general spam sample. Extracting these attributes from other email corpora may result in rather sparse data. Fortunately, a special spam track within the context of the larger Text REtrieval Conference (TREC), a U.S.-government-supported program that facilitates analysis, evaluates participants' filters on real email streams (Goodman et al., 2007). In addition, standard measures and corpora are defined for tests in the future. The spam track depends on two types of email corpora. The first is synthetic, made up of a rare public corpus of non-spam messages and combined with a carefully modified set of recent spam. Researchers run their filters on it and it may be freely shared. With the second private corpora, researchers submit their code to testers who run it on the corpora and return summary results only, thereby guaranteeing privacy.

ACKNOWLEDGMENTS

The authors wish to thank SAS® Institute for providing them with Base SAS® and SAS® Enterprise Miner™ software used in computing all the results presented in this paper. This work forms part of research done at the North-West University within the TELKOM CoE research program, funded by TELKOM, GRINTEK TELECOM and THRIP.

REFERENCES

Asuncion, A. and Newman, D. J. (2007), UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences, http://www.ics.uci.edu/_mlearn/MLRepository.html, Date of access: 17 February 2010.

Berk, K. N. and Booth, D. E. (1995), 'Seeing a curve in multiple regression', *Technometrics* 37(4), 385–398.

Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford.

Cranor, L. F. and LaMacchia, B. A. (1998), 'Spam!', *Communications of the ACM* 41(8), 74–83.

De Waal, D. A. and Du Toit, J. V. (2007),Generalized additive models from a neural network perspective, *in* 'Workshops Proceedings of the 7th IEEE International Conference on Data Mining (ICDM 2007), October 28-31, 2007, Omaha, Nebraska, USA', IEEE Computer Society.

Du Toit, J. V. (2006), Automated Construction of Generalized Additive Neural Networks for Predictive Data Mining, PhD thesis, School for Computer, Statistical and Mathematical Sciences, North-West University, South Africa.

Ezekiel, M. (1924), 'A method for handling curvilinear correlation for any number of variables', *Journal of the American Statistical Association* 19(148), 431–453.

Goodman, J., Cormack, G. V. and Heckerman, D. (2007), 'Spam and the ongoing battle for the inbox', *Communications of the ACM* 50(2), 25–33.

Hastie, T. J. and Tibshirani, R. J. (1986), 'Generalized additive models', *Statistical Science* 1(3), 297–318.

Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized Additive Models*, Vol. 43 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, London.

Kiran, R. and Atmosukarto, I. (n.d.), Spam nor not spam – that is the question.

Larsen, W. A. and McCleary, S. J. (1972), 'The use of partial residual plots in regression analysis', *Technometrics* 14(3), 781–790.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, Vol. 37 of *Monographs on Statistics and Applied Probability*, 2nd edn, Chapman and Hall, London.

Potts, W. J. E. (1999), Generalized additive neural networks, *in* 'KDD '99: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, New York, NY, USA, pp. 194–200.

Ripley, B. D. (1996), *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge, United Kingdom.

Wood, S. N. (2006), *Generalized Additive Models: An introduction with R*, Texts in Statistical Science, Chapman & Hall/CRC, London.

Zhang, G., Patuwo, B. E. and Hu, M. Y. (1998), 'Forecasting with artificial neural networks: The state of the art', *International Journal of Forecasting* 14, 35–62.

OPERATIONS RESEARCH SOCIETY OF SOUTH AFRICA

2011 STUDENT COMPETITION

Project work that was undertaken for a Masters or Honours in Operations Research or a related field of study during the **2010 academic year,** may be entered. Only projects undertaken by individuals will be considered.

Objectives:

- to propagate the use of Operations Research (OR),
- to encourage the inclusion of project work in courses within the field of OR,
- and to bring the Operations Research Society of South Africa (ORSSA) to the attention of students and staff at universities and technikons.

Please contact the organizer for **application forms** and important **closing dates**.

Margarete Bester contactable at,

mbester@oprecon.com

DISCLAIMER

HARNESSING THE POWER OF DATA TO OPTIMISE BUSINESS RESULTS

Francois Beyleveld

Francois Beyleveld, at SAS, explains why sustainability as a concept offers a watershed of opportunity for better business performance through innovation, while also benefiting the planet and employees' own careers.

More and more South African companies are beginning to realise that 'greening' their IT infrastructures brings greater business efficiency, return on investment and improved levels of service to their organisations.

In fact, organisational performance as we know it, demands sustainability measures across social, environmental and economic factors, which in turn requires the vital steps of integrating and analysing data to achieve new goals and transform internal organisational cultures. Technology companies in particular are at the forefront of green IT initiatives, because they acknowledge that their reputation as socially responsible entities is critical.

The right choices

By deploying the right technologies, IT can play a significant role in furthering a company's ability to monitor, analyse and implement more sustainable, or green practices, defined as those that meet the requirements of the present day without compromising the ability of future generations to meet their needs. In many cases, making small, incremental changes in IT processes can lead to definitive benefits.

When it comes down to the practicalities of doing business in an increasingly energy-hungry world, most companies are now looking at their supply chain and their ability to measure, monitor and improve their efficiency footprint within their organisations. Those that cannot demonstrate that value, might find themselves out in the cold when the next tender request comes around.

Another aspect that is becoming more important in green IT is employee retention and recruitment. Talented employees have many employment options and are increasingly looking at their employers and their stance on corporate responsibility and the environment. Clearly, reducing energy use is more about being environmentally responsible. To ensure their long-term viability, organisations must begin now to find and implement solutions that decrease power consumption.

The tools

The good news? Many of the same tools and practices that have enabled these organisations to reduce IT complexity, streamline operations and controls are also highly effective in energy use to help companies become lean, clean and green organisations. These include data de-duplication, high availability and virtualisation, power management and energy efficient data centre design.

The most strategic enterprises will use data, and the intelligence gained from it, to their competitive advantage – driving increased brand value through innovation and improving internal efficiencies and accountability. They will also build loyalty in consumers, employees and other stakeholders – such as in higher education where they track, communicate and educate on sustainability.

Today, companies are able to measure, manage and report on the Triple Bottom Line – environmental, social and economic indicators – and determine business strategies to reduce risk and increase shareholder value.

The results

Harnessing sophisticated software, companies are able to measure key sustainability activities using methodologies and protocols, utilising their existing data in operational systems and databases.

They are also able to report ongoing performance to ensure transparency with key stakeholders and compliance with regulatory agencies. By establishing an integrated, consistent source of quality information, companies can bind initiatives to a common sustainability framework that allows alignment across all lines of business – from water treatment facilities to the data centre,

Additionally, companies are able to improve performance by identifying metrics that have the greatest impact on goal attainment so that they can make the most informed strategic decisions by using optimisation, forecasting and data mining capabilities to analyse scenarios and run simulations to improve response and successful strategy execution.

Organisations can also manage and forecast the finances and resources needed to achieve the desired outcomes across the enterprise and within each department. Using analytics, they are able to prioritise organisational strategies and align investments in new product innovation, programme management and talent accordingly and establish scorecards and strategy maps driven by the sustainability goals of the organisation.

To end

In closing, going green offers a vital path to innovation and creating enduring value and competitive advantage. Despite the challenges of adopting an environmental mind-set, the direction that companies have to head in is clear, and it is clear that IT has a key role to play. When people start understanding the strategic risk and strategic opportunities of climate change in terms of its impact on brand value, their market and their operations, they'll get engaged in a much broader environmental agenda.

To learn more about how to meet the requirements for real-time decision making, contact SAS on +27 11 713 3400 (Johannesburg and Pretoria) or +27 21 912 2420 (Cape Town) or visit www.sas.com/sa

MEMBER INTERVIEW: WINNIE PELSER

Contactable at: winniep@armscor.co.za

Winnie Pelser

My first degree was in Mathematics and Statistics. My husband was transferred overseas and I exploited the opportunity for formal further studies in OR through UNISA. It was Before Internet and distance studies were a real challenge. Electronic banking and e-mail was then only mentioned in science fiction books. Prof Uri Passy of Haifa Technikon served as promoter for my Masters thesis.

After our travels I worked on modelling at Denel Simulation and Wargaming and got involved in software quality, which necessitated some further study at the University of the Witwatersrand and qualifying as a Certified Quality Analyst through the US Quality Assurance Institute's software/ systems engineering and management course. This was extremely interesting and very much in line with OR since the approach was definitely "holistic". The section was however closed and after a stint in software quality in the private industry I ended up at Armscor doing Defence OR again. I am currently a decision support analyst at Armscor's Defence Decision Support Institute.

I am married to Johan, an officer in the South African Air Force. Our daughter Anro is studying architecture and our son Ian, engineering.

When and how did you first become involved in OR and ORSSA?

My first employment was with Armscor's Logistics and *Operations Research* (OR) division. Our manager was Dr Jos Grobbelaar, who lectured in OR at UNISA part time. He got me interested. Jos was involved in ORSSA for many years. I was very lucky to come across the field so early, and it has fascinated me ever since.

You practice OR in a military environment. Can you give us any information on what your work entails in a military environment?

The institute delivers a decision support capability to the South African National Defence Force. We also constantly research the application of tools and methods in the defence environment. As with all research it is essentially exploratory – there is no guarantee of success. Often the impact may only be realised in the longer term, in conjunction with other work and after several iterations.

A model of decision-making proposed by Malaska and Holstius[1] lists three kinds of knowledge as input to sound decisions:

- Knowledge about purpose and objectives;
- Situational knowledge; and
- Knowledge about means and resources.

These are fused with the wide variety of available supporting methodologies, techniques and tools to arrive at good decisions through perception and sound logic. Defence decision support analysts must possess:

- Technological excellence and scientific creativity;
- Independence and impartiality;
- Trusted access to the military;
- Defence domain knowledge and experience.

Supported decisions typically have significant and persistent implications. Although accountability remains with the decision maker, the quality of analysis results and emanating recommendations presented by OR teams is critical.

You have done a lot of work on soft system methodology. Can you give some clarity on the difference between soft systems and more hard, mathematical techniques, and in what context each of these can add any value?

Higher-level decisions often involve complex systems, which may be defined as, "system[s] that [are] comprised of a large number of entities that display a high level of interactivity"[2]. This interactivity is often non-linear and contains feedback loops. Complex

[1] **Pentti Malaska and Karin Holstius. Visionary Management in *Foresight, Vol.01, no.04, Aug.99***

decisions thus involve both quantitative and/or qualitative data, with complex and often obscure inter-relationships. Recent international trends in decision science include:

- The recognition of complexity in (high-level) decision making.
- The acceptance of the validity of "Hard" and "Soft" techniques and their applicability to different types of problems, or even aspects of the same problem.
- The growing importance of "sense-making" and of gaining understanding in a complex environment, as opposed to seeking a single correct or optimal answer.

Soft systems methodology, the result of research by Peter Checkland and others, is a systemic approach for tackling problematic situations. It provides a framework for handling ill-defined or not easily quantified (in other words, messy) problems that often do not have a commonly agreed set of outcomes. Traditional systems analysis is not very appropriate for dealing with such problems. Soft system analysis attempts to understand complexity, promote learning, to identify weaknesses and to understand relationships.

In hard systems approaches rigid methods are used to provide unambiguous solutions to well-defined data and processing problems. Hard systems approaches assume that problems are well defined, that the traditional scientific approach to problem solving will work and that technical factors will dominate.

With hard OR there is a single decision maker with a clear objective. Soft OR may involve a range of decision makers or groups with conflicting objectives. The most important factors can be quantified and it is possible to collect data with hard OR. With soft OR it is often not possible to quantify the most important factors.

Hard OR practitioners require good analytical skills and fulfil the role of hard analyst. The role of the soft OR practitioners is one of facilitator. The important lesson is that one approach is not better than the other. It is important to use the appropriate method or approach for the situation at hand. The interesting issue is whether it is possible to use a combination of the two in a single application, depending on the context of the problem.

What have been the highlights of your OR career?

I was involved in a software development process that was put in place from the very start all the way to testing according to best practices and Military specification for the private sector. It was an interesting and satisfactory opportunity.

I cannot elaborate about the real Military work I was and am involved in for security reasons. But every time when a project you worked on is successfully implemented or if you helped solve or clarify a difficult decision it is a highlight.

Thanks to a very enthusiastic group our involvement in the ORSSA chapter proved to be very successful. Our aim was to 'market' OR as wide as possible. So far we reached a number of people and hope it will lead to more involvement.

What aspect of Operations Research do you fancy the most and why?

The one aspect of OR that keeps me fascinated is the variety of methods and thought processed in the field. It is stimulating and you can learn something new continuously!

The United States Air Force academy includes operations research in their curricula. According to James K. Lowe and Col. Andrew P. Armacost[3], it is one of the most popular courses at the Air Force Academy. As they aptly said, they will not make the diverse group of students all OR analysts, but the class appreciates the possibilities and existence of the field of OR. One of their students quoted: "This O.R. stuff will fundamentally change the way I view the world!"

Do you have a message for the young aspiring OR students/practitioners out there?

This is a wonderful field for both research and application. The possibilities are endless. Apply well known methods and approaches, but keep on investigating new ones. Remember to make sure you apply appropriate solutions to problems. May "this OR stuff change the way you think about the world".

[2] **Richardson, Kurt A. , Graham Mathieson and Paul Cilliers, *The Theory and Practice of Complexity Science: Epistemological Considerations for Military Operational Analysis,* Available from: www.kurtrichardson.com/milcomplexity.pdf.**

[3] **James K. Lowe and Col. Andrew P. Armacost, United States Air Force Academy. 'All for one and 'O.R. for all', August 2010 OR/MS Today**.

A new online submission system for ORiON

by Martin Kidd (14623617@sun.ac.za) and Stephan Visage (svisagie@sun.ac.za)
Department of Logistics, University of Stellenbosch

Since the beginning of 2010 we have been working on implementing an online submission system for ORiON with the help of the *Open Journal Systems* (OJS) software, freely available online from the *Public Knowledge Project* (PKP) website (pkp.sfu.ca). This system is aimed at streamlining the submission, reviewing and editing processes, to facilitate the record keeping of submission history, e-mail correspondence and referees' reports, and to provide an online basis for authors to follow the progression of their submissions. The purpose of this newsletter article is twofold, namely to (1) invite operations researchers to utilise this submission system when submitting work to ORiON, and (2) to highlight some of the system's primary features.

Figure 5 shows the system's home page. If you are a registered user, you may log in on this page by entering your username and password in the provided spaces (simply click on "Log in" if you have forgotten your password). If you are not yet registered with the system, you may click on the link "Register" at the top of the page, and complete the registration form. There are some other features on the home page which are available to non-registered users (or users not logged in). First of all, archives of previously published papers in ORiON are available, and on the left hand side of the home page, search facilities are provided for the purpose of browsing for particular papers. For general browsing the archives may alternatively be accessed via the link "Archives" at the top of the page. Another feature available to non-registered users may be accessed by following the "About" link at the top of the page. Here you may find useful information regarding the editorial team, the editorial and advisory boards, ORiON's review and publication policies, submission guidelines, and more.

Once logged in, clicking on "User home" takes you to the user homepage, shown in Figure 6. On this page your current roles with ORiON are listed, which is either author, reviewer, or both. Clicking on a specific role will take you to a page listing all manuscripts that either you have submitted as author, or that you have been

Figure 5: ORiON online submission system home page.

assigned to review. The user homepage also provides links to editing your profile details and changing your password. If you follow the "Author" link, you may submit a new manuscript by clicking on the link below "Start a new submission" at the bottom of the page.

If you have already submitted manuscripts to ORiON as an author, you will find them listed as either active or archived. Clicking on the title or status of a manuscript takes you to the submission page which contains three further links, namely "summary", "review" and "editing". "Summary" contains information that you have entered as author when first submitting the manuscript, as well as the manuscript status, whereas "review" contains referees' reports (once completed) and the editor's decision on your manuscript (once made). The "editing" page is only used by the author if the manuscript is eventually accepted for publication, and only to check the final proofs of the manuscript before publication.

Following the "Reviewer" link on the user homepage you will find a list of manuscripts you are currently reviewing. Clicking on the name of a manuscript takes you to a page containing the title and abstract of the manuscript, the review schedule, and the review steps. The review steps include either accepting or rejecting the invitation, downloading the manuscript, completing the provided online review form, uploading any additional files (such as a list of corrections and/or comments), and finally giving your overall recommendation. These steps need not be completed all at once, you may revisit the webpage at your own leisure in order to complete the steps.

The submission system has undergone a testing phase, and we would like to extend a word of thanks to all authors and reviewers who assisted us in this task. After clearing up a small number of hiccups, we are happy to report that the system will officially be put to use from 2011 onwards. Since the OJS software is updated every now and then, some minor bugs or other difficulties may present themselves in time, and it would be much appreciated if these problems could be reported to the journal manager, Martin Kidd (orion.oss.support@gmail.com), once picked up by a user. You may also use this email address for support if you find yourself stuck at any point during the submission or review processes. Finally, we invite researchers to no longer submit their manuscripts via e-mail to the editor as was done in the past, but rather to visit ORiON's online submission system (to which a link will be provided via the ORSSA website) and to submit their manuscripts online. We hope that this submission system will prove to be as convenient and beneficial for authors and reviewers as it has been for the ORiON editorial team.

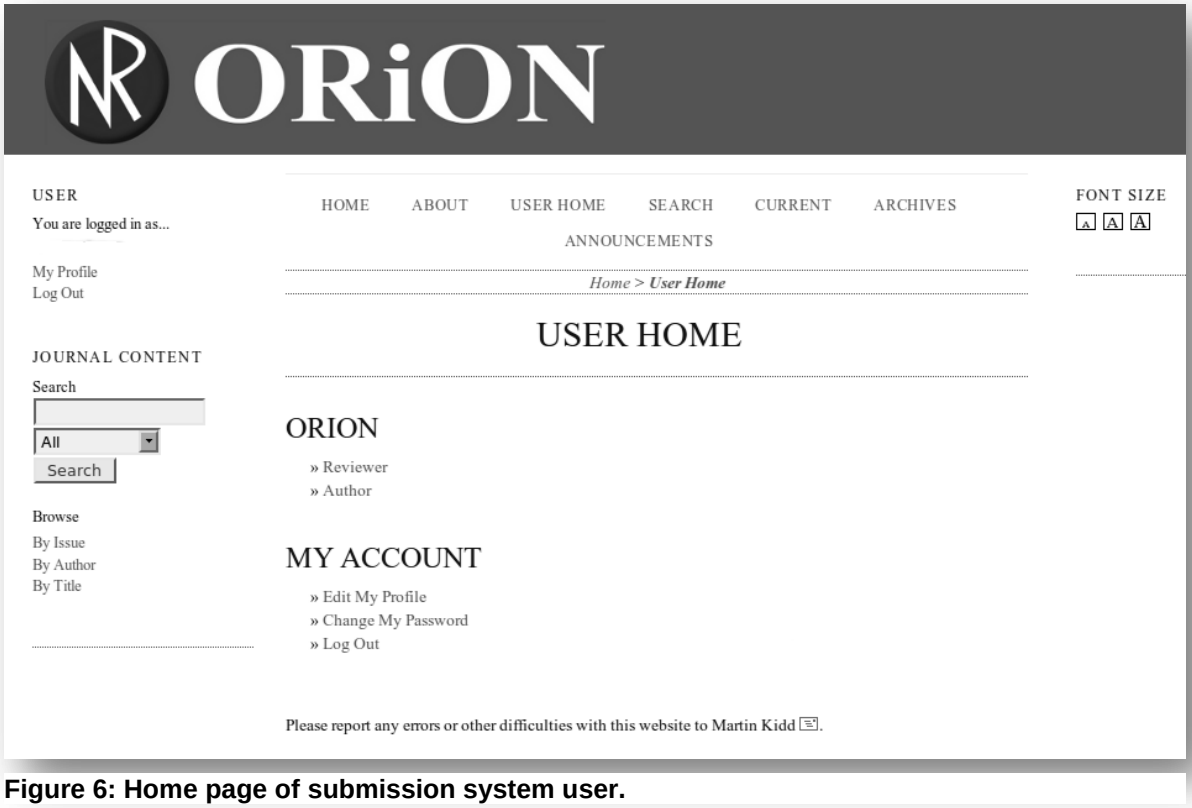


Figure 6: Home page of submission system user.

The Tom Rozwadowski award: Nominations now open

The Tom Rozwadowski medal is the Society's premier award and has been awarded almost every year since its inception in 1971. The medal is awarded for the best paper published in a local or international peer-reviewed Operations Research journal by a member of the Society during the previous year; 2010, in this case. (The editor of ORiON will automatically nominate all papers published in ORiON by members of the Society during 2010.)

The nominating committee invites submissions for consideration for this award. Nominations should be submitted to the chairman of the nominating committee, the ORSSA Vice-President, Jan van Vuuren, (vuuren@sun.ac.za).

The closing date is 1st June 2011.

The following rules apply for the Tom Rozwadowski award:

1. Contributions of an OR nature published in journals of international standing during the previous year, are eligible for consideration.
2. Confidential or secret material will not be accepted for consideration.
3. Only persons who were members of the Society, or who had already applied to become members of the Society when the contribution was made, are eligible for the award.
4. Contributions will be screened by the nomination committee (consisting of the vice-president [convenor], the chapter chairpersons and the archivist) and adjudicated by a selection committee (consisting of the president, the vice-president and two members of the executive committee), which will only consider material submitted by the nominating committee.
5. Any member of ORSSA may submit a contribution for consideration or draw it to the attention of the nominating committee, whether they are an author or not.
6. The nominating committee shall submit at least two contributions to the selection committee.
7. The selection committee may appoint expert referees for all of the contributions under consideration.
8. Should a member of the selection committee be under consideration for the award, he/she shall excuse him/herself, and a replacement member shall be co-opted to the selection committee by the members of that committee.
9. Where the winning material was produced by co-authors, every co-author who is a member of the Society shall receive a medal.
10. One or more of the following criteria, inter alia, will be used as a basis for making the award:
    - 10.1. Originality
    - 10.2. The quality of any theory developed
    - 10.3. Interaction between theory & practice
    - 10.4. New areas of application
    - 10.5. New opportunities created for Operations Research
    - 10.6. Clarity of exposition.
11. Contributions should preferably be in English.
12. Members are encouraged to participate and the chapter chairpersons, in particular, are requested to ensure that all worthy material originating in their region is brought to the attention of the nominating committee.

OPERATIONS RESEARCH SOCIETY OF SOUTH AFRICA

OPERASIONELE NAVORSINGSVERENIGING VAN SUID-AFRIKA

2011 Annual Conference: Spreading Operations Research Across Africa

Call for papers

The 40th Annual Conference of the Operations Research Society of South Africa (ORSSA) will be hosted by the National University of Science and Technology, Bulawayo, Zimbabwe. The Conference will take place from 18th to 21st September 2011, with 22nd September reserved for sightseeing. It will be held at the Elephant Hills Hotel at Victoria Falls, Zimbabwe. In holding the conference in Zimbabwe, ORSSA hopes to foster greater collaboration and cooperation across southern African OR communities, accelerate the expansion of OR applications and education in Zimbabwe, and encourage the establishment of a Zimbabwean OR society.

With the theme of **"Spreading Operations Research Across Africa,"** ORSSA is moving in its stated direction and working to expand the awareness, understanding, and use of OR across the wider region. This theme encourages participation over the full spectrum of Operations Research, welcoming papers of a more fundamental nature, those on the application of Operations Research techniques in business and industry, about topical issues in Operations Research, and about the philosophy, teaching and marketing of Operations Research.

Those interested in participating in the Conference should submit an abstract of no more than 300 words via the website below with effect from 14th March 2011 (by clicking on "Delegates" and then on "Abstract Submission"). The abstract should contain plain, unformatted text, with no mathematical expressions or formulae. The deadline for admissions is 29th July 2011, and notification of acceptance will be given via email by 22nd August 2011.

Prospective delegates may register for the Conference via the website below from 14th March 2011 (by clicking on "Delegates" and then on "Registration"). The deadline for early-bird registration is 11th July 2011, and the last day of registration is 16th September 2011.

Delegates are responsible for their own travel and accommodation arrangements. The Elephant Hills Hotel is strongly recommended, as the Society has arranged discounted rates for delegates. ORSSA is negotiating for reduced costs for airfares, and anyone who plans to fly to Victoria Falls and wishes to take advantage of these should record that in the relevant box on the website registration.

As anyone who has attended previous ORSSA conferences will be aware, the social side of the conference is also important. We recommend the Elephant Hills Hotel as the preferred place to stay, to allow full involvement in all the activities around the conference.

Registration discounts are available for members of ORSSA, for early-bird registration and for students.

A preliminary overview of the structure of the conference may be found on this website by clicking on "Programme" and then on "Overview".

Please visit the conference website for information:

www.orssaconf.org.za

IFORS 2011
Conference for the International Federation of Operational Research Societies

Second Call for Papers
19th Triennial Conference of the International Federation of Operational Research Societies
10th – 15th July, 2011

Melbourne, Australia

World OR: Global Economy and Sustainable Environment

The 19th Triennial Conference of the International Federation of Operational Research Societies (IFORS) will be hosted by the Victorian chapter of the Australian Society for Operations Research (ASOR). The conference will be held at the new Melbourne Convention Centre in the centre of the city of Melbourne and will bring operational researchers from around the globe together.

Abstract Submission:

Papers on all aspects of Operational Research are invited. Authors wishing to present are requested to submit an abstract of not more than 100 words via the Abstract Submission system at: http://www.euro-online.org/conf/ifors2011/
Please visit the webpage for further information:
http://www.ifors2011.org

ECCO XXIV
European Chapter on Combinatorial Optimization

ECCO XXIV
24th Annual Meeting of European Chapter on Combinatorial Optimization (ECCO)
May 30th – June 1st, 2011, Amsterdam, The Netherlands

The ECCO annual meeting aims to bring together researchers in the field of Combinatorial Optimization to present their work, share experiences, and discuss recent advances in theory and applications. The primary objectives are:

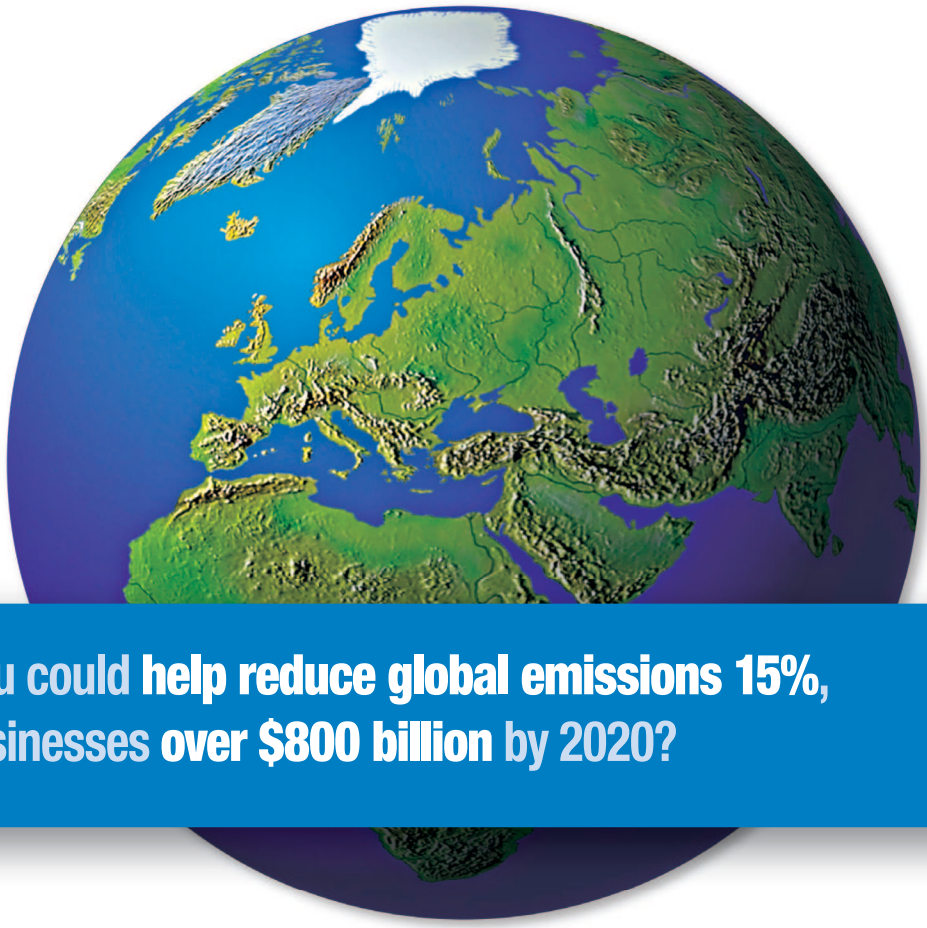
- exchanging results and experiences in solving real-world combinatorial optimization problems,
- reporting on development and implementation of appropriate models and efficient solution methods for combinatorial optimization problems,
- establishing networking contacts between individuals and research groups working on related topics,
- promoting the work on combinatorial optimization (theory and applications) to the broader scientific community,
- identifying challenging research problems for the field, as well as promising research outlets (both in theory and applications),
- promoting interactions with researchers in other related fields.

This conference provides an excellent opportunity to discuss recent and important issues in Combinatorial Optimization and its applications with European combinatorialists; most European countries are represented.

http://www.eccoxxiv.com

SAS® for Sustainability Management

Predictive Analytics | Performance Reporting | Carbon Footprint Modeling | Data Management
Resource Optimisation



What if you could **help reduce global emissions 15%,** saving businesses **over $800 billion** by 2020?

You can. SAS gives you The Power to Know.®

SAS for Sustainability Management enables your organisation to measure, manage and report on environmental, social and economic indicators, as well as develop business strategies that reduce risk and increase brand value.

▸▸ **www.sas.com/sa**
for a free white paper

sas | THE POWER TO KNOW.