HSE
Health & Safety
Executive

Industrial use of safety-related artificial neural networks

Prepared by
Liverpool John Moores University
for the Health and Safety Executive

CONTRACT RESEARCH REPORT

327/2001

HSE
Health & Safety
Executive

Industrial use of safety-related artificial neural networks

Professor Paulo J G Lisboa
Liverpool John Moores University
School of Computing and
Mathematical Sciences
Byrom Street
Liverpool
L3 3AF

The overall objective of this study is to investigate to what extent neural networks are used, and are likely to be used in the near future, in safety-related applications.

Neural network products are actively being marketed and some are routinely used in safety-related areas, including cancer screening and fire detection in office blocks. Some are medical devices already certified by the Food and Drug Administration (FDA). The commercial potential for this technology is evident from the extent of industry-led research, and safety benefits will arise. In the process industries, for instance, there is real potential for closer plant surveillance and consequently productive maintenance, including plant life extension.

It is clear from the applications reviewed that the key to successful transfer of neural networks to the marketplace is successful integration with routine practice, rather than optimisation for the idealised environments where much of the current development effort takes place. This requires the ability to evaluate their empirically derived response using structured domain knowledge, as well as performance testing. In controller design, the scalability of solutions to production models, and the need to maintain safe and efficient operation under plant wear, have led to the integration of linear design methods with neural network architectures.

Further research is necessary in two directions, first to systematise current best practice in the design of a wide range of quite different neural computing software models and hardware systems, then to formulate a unified perspective of high-complexity computation in safety-related applications.

There is a need to develop guidelines for good practice, to educate non-specialist users and inform what is already a wide base of practitioners. Combined with a safety awareness initiative, this would be of as much of benefit to the development of this commercially important new technology, as to its safe use in safety-related applications.

This report and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the author alone and do not necessarily reflect HSE policy.

HSE BOOKS

First published 2001

ISBN 0 7176 1971 0

CONTENTS

1 OBJECTIVES AND SUMMARY

The overall objective of this study is to investigate to what extent neural networks are used, and are likely to be used in the near future, in safety-related applications. In detail:

- To describe and categorise current industrial safety-related applications of neural networks.
- To discuss the issues arising from current industrial use of neural networks, including: potential safety gains; difficulties in achieving and demonstrating safety; identifiable trends.
- To indicate other key centres of excellence in neural networks and their application.

Neural network products are actively being marketed and some are routinely used in safety-related areas, including cancer screening and fire detection in office blocks. This includes medical devices already certified by the FDA.

The commercial potential for this technology is evident from the extent of industry led research. Yet, the perception from an evaluation report of the DTI funded Neural Computing Technology Transfer Programme involving 3,500 companies was that, in 1998 'neural computing had not yet reached a stage where it could be demonstrably shown to be effective' [9]. To this most practitioners would add that it is not perceived to be demonstrably safe, either. Nevertheless, safety benefits will arise. In the process industries, for instance, there is real potential for closer plant surveillance and consequently productive maintenance, including plant life extension.

Good practice was observed in the successful commercial applications reviewed in this report. It is clear from the applications reviewed that the key to successful transfer of neural networks to the marketplace is successful integration with routine practice, rather than optimisation for the idealised environments where much of the current development effort takes place. This requires the ability to evaluate their empirically derived response using structured domain knowledge, as well as performance testing. In controller design, the scalability of solutions to production models, and the need to maintain safe and efficient operation under plant wear, have led to the integration of linear design methods with neural network architectures

Further research is necessary in two directions. The first is to systematise current best practice in the design of a wide range of quite different neural computing systems, some for static pattern recognition, others for control, some in software and others embedded in hardware. The second is to formulate a unified perspective of each stage in the development lifecycle of high-complexity computation in safety-related applications, spanning the full range from empirical to structural process models. In emerging computation, the complexity is often not in the software implementation, but in the interpretation and testing required to evaluate the operation of the model.

There is an opportunity for voluntary regulation, which may be implemented by issuing guidelines for good practice, to educate non-specialist users and inform what is already a wide base of practitioners. This could be backed-up by the award of quality seals to encourage best practice as a route to commercial advantage, with specific levels of compliance including a commitment to report adverse events. Combined with a safety awareness initiative, this would be of as much of benefit to the development of this commercially important new technology, as to its safe use in safety-related applications.

2 KEY TECHNICAL CONCEPTS

2.1 THE NEURAL NETWORK PARADIGM

Artificial neural networks share their origins with the infancy of machine-based information processing, when McCulloch and Pitts first showed that a network of interconnecting threshold units can replicate any Boolean function. These units are modelled on the response of neural cells in biological nervous systems, hence the evocative name given to this field. Later, Rosenblatt carried out his investigations with analogue electro-mechanical systems for visual pattern recognition, developing the seminal concept of the perceptron as an embodiment of the McCulloch-Pitts already idealised model of the neuron. This work demonstrated a key property of artificial neural systems, namely their distributed associative memory function. This means to say that the new information is stored in the weighed links between the threshold units, rather that at specified memory addresses. An immediate consequence of this is that the component elements of the stored memories, for instance on-bits in a binary image, do not have a one-to-one correspondence with any network parameters. Instead, the overall recall accuracy degrades slowly as weight-links are disturbed, rather like seeing through frosted glass. A physical example of this phenomenon is holography, where the resolution of the reconstructed 3-D image increases with the size of the film used to capture the interference pattern that makes up the hologram.

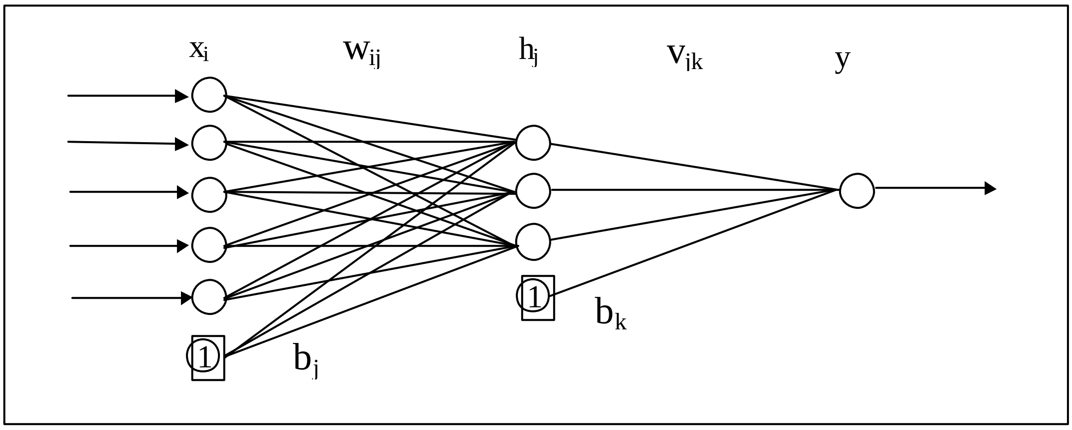

Figure 1
Artificial neural network architecture, whose key feature is the representation created by an intermediate, or hidden layer of processing units with non-linear transfer functions. This schematic represents the feed-forward path for the Multi-Layer Perceptron (MLP): *x, h* and *y* are the layers of neurons, *w* and *b* are the weights and bias terms.

Having carried out his doctorate work on the perceptron, Minsky proceeded to discredit the whole field by pointing out that a single layer of adaptive link-weights could only resolve linearly separable patterns, failing on pattern recognition patterns as simple as the exclusive-OR. It was already pointed out at this time that multiple layers of weights would overcome this difficulty, but it was naturally assumed that the computational difficulties would simply transfer into the optimisation of the network parameters, which would get trapped by a proliferation of local minima.

The current standing of this technology owes much to the affordability of fast computation, when Rumelhart and collaborators popularised the multi-layer perceptron (MLP) as a practical paradigm for distributed associative memory. They demonstrated that the MLP is a generic 'learning' machine capable of forming representations of complex, non-linear phenomena. Local minima turn out to be saddle points in all but the smaller networks, as the number of escape routes increases exponentially with the dimensionality of weight space.

The widespread uptake of neural networks in the late 80's had two contrasting outcomes. It resulted in naïve applications that would not usually merit publication in statistical or engineering journals, and lent the field a bad name. However, research effort generated by this popularity also resulted in important theoretical results, starting with the demonstration that MLPs were indeed universal approximators for non-linear regression and classification tasks. Critically, this flexibility is also the main weakness of this computing paradigm, by causing overfitting of the data, that is to say modelling noise effects. This led to a 'second generation' of neural models where statistical approximations are used to 'regularise' the networks, using self-consistency to protect against overfitting.

The occurrence of overfitting is a natural consequence of non-linear data modelling, represented in fig. 2 by two alternative noise models to fit a small size data sample. The model is a Gaussian Process (GP), which predicts statistically valid non-linear error bars, but each solution is a different local minimum of an objective function with two terms, one to promote accurate fitting of the data, and the other to penalise model complexity.

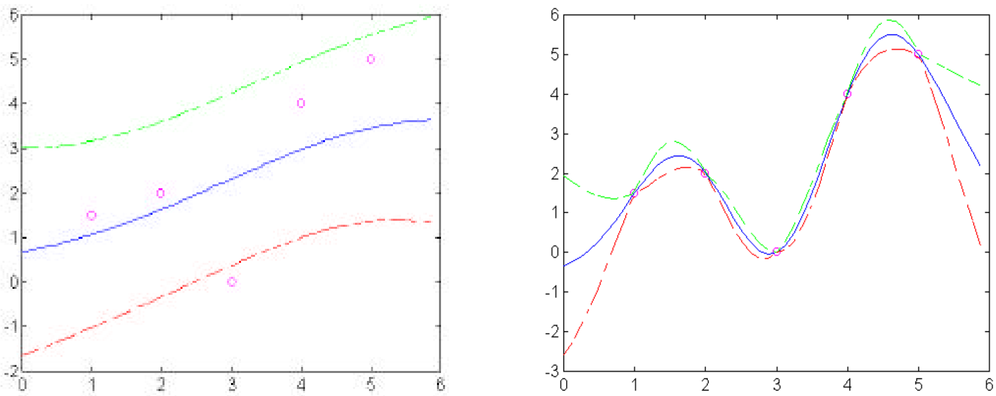

Figure 2
Which is the best model to apply to future data? Although it is illustrated here with respect to non-linear regression in a continuous domain (http://wol.ra.phy.cam.ac.uk/mackay/abstracts/gpB.html), the same effect applies to classification tasks where class boundaries can also be unnecessarily complex. The difficulty in separating useful from redundant information is common to all inference systems.

2.2 PERSPECTIVE OF INDUSTRIAL APPLICATIONS

The Annual Report to the US Securities and Exchange Commission of Neuromedical Systems Inc., for the fiscal year to December 1997 describes neural networks as

> 'representing a relatively new computer technology. Instead of the single path of programming utilised by conventional computers, neural network computers have thousands of adaptively-formed paths linked in parallel, comprising a network conceptually similar to the network of neurons in the human brain. Neural networks can process immense quantities of information while mimicking the neurological ability to learn from experience. After training on images of a series of "abnormal" and "normal" cells, a neural network can recognise abnormal cells even though they differ

> considerably from the training set of images. Neural networks excel at solving complex pattern-recognition problems'.

The exaggerated perception that neural networks emulate 'little brains' is both a precursor for the resurgence of neural networks in the '80s, and the source of a plethora of ill-defined practical applications - some of which, against the odds, turned out to be successful.

A characteristic feature of this technology is that it spawned useful commercial products from the earliest days of non-linear modelling. Among five well known early products, four of them are still successful a decade later. These early products are listed below:

Airline bookings

BehavHeuristics, Inc., started in 1986 and now part of Airline Automation, Inc. (http://www.airauto.com/aai/), use reinforcement learning to predict no-shows in air flights, thus maximising the passenger load through controlled overbooking [1]. Their Airline Marketing Tactician (AMT) was an early success for neural networks and remains a market leader today.

Design retrieval system

Boeing Company's NIRS (Neural Information Retrieval System), is probably still the largest-scale manufacturing application of neural networks. It uses a binary Adaptive Resonance Theory network (ART1), to cluster binary templates of aeroplane parts in a complex hierarchical network that cover over 100,000 items, grouped into thousands of self-organised clusters. Claimed savings in manufacturing costs are in millions of dollars per annum.

Explosive detection

SNOOPE from SAIC [1], is an explosive detector that was deployed in several high-risk airports including Heathrow and Los Angeles, from around 1987, motivated by the need to detect the plastic explosive Semtex. The detector worked by irradiating suitcases with thermal (low energy) neutrons, collecting an emission gamma-ray spectrum and classifying this into three categories, namely bulk explosive, sheet explosive, and no explosive. A standard MLP was benchmarked against multivariate linear regression, and found to have a similar area under `the Receiver Operating Characteristic (ROC) curve, which implies equivalent discriminatory power, but with an improved sensitivity of 92% at a specificity of 96%. This was the target false-alarm rate, hence the selection of the MLP for this product. However, there were severe practical difficulties with this product, primarily the bulk of lead shielding required, as well as the sheer volume of passenger traffic which meant that even 4% of false positives represent a large number of items to check by alternative means, quickly reducing the return on investment.

Cytological screening

Neuromedical Systems, Inc., were granted US patent 4,965,725 for a 'Neural network based automated cytological specimen classification system and method' in 1990, two years after filing for it. This system was commercialised under the name of Papnet. It is among the most widely tested medical decision support systems to-date, and leads the use of any form of artificial intelligence [2]. To quote from the 1997 Annual Report,

> 'the FDA pre-market approval application involved a study of more than 10,000 Pap smears originally classified as "negative" and retrieved from laboratory archives. The trial measured the ability of Papnet testing to assist in the detection of missed abnormalities on Pap smears previously diagnosed as "negative" from a population of women who nevertheless subsequently developed high grade lesions or invasive

cervical cancer (the "Case" group). The system was also used to re-examine "negative" Pap smears from arbitrarily selected women who were not known to have developed cervical lesions or cancer (the "Control" group). The trial results indicated that, when used as an adjunct to manual screening, Papnet testing can increase the aggregate cervical abnormality detected by up to 30% when compared with the combination of manual screening and routine manual quality control re-screening. The trial results also demonstrated that Papnet testing assisted in the detection of abnormality which had been originally missed by manual screening for 31.6% of the Case group. For 91.7% of such women, Papnet testing would have found the abnormality more than a year prior to the biopsy that confirmed the patient's disease, and, for 38.9% of such women, more than two years earlier. In 1996, the Company received an additional claim from the FDA that shows that Papnet testing of routine presumed negative Pap smears can be conservatively expected to identify seven times more false negatives when compared with manual quality control re-screening of the same number of smears.'

Further to this, a detailed study of its clinical effectiveness involving 200,000 Pap smears estimated a standard measure of cost effectiveness, the cost per life-year saved, to be US$48,474 at 1994 prices, for the addition of Papnet re-screening into a routine screening protocol on a biennial basis. This is in contrast with $67,918 estimated for annual mammography screening of women aged 40-49, and $113,000 for one test for Prostate Specific Antigen. Yet, even at a modest additional cost per test to the patient, cost-benefit and other considerations have caused a protracted wait to capture a sizeable proportion of the market, eventually leading the original developer, to register trading losses of $131,089,000 by the end of 1997, followed by a take-over by the rival company NeoPath, Inc (http://www.neopath.com/).

This case study illustrates the demands for cost-efficiency now driving much of the healthcare market, which exacerbate the already substantial costs of certification. Similar considerations apply to other safety-related fields, notably in the process industries. The profit to be made from the development of intelligent systems in niche markets, even comparatively large ones such cancer screening, may be quickly off-set by development and regulation costs, especially when the innovation agent is a small company. Yet SMEs form the bulk of healthcare care industries both in the UK, as in the US. When products are highly customised, or address niche markets, even effective neural network-based safety technology may prove difficult to introduce to industry.

Financial risk management

The other neural network applications noted for their early transfer from the laboratory to the marketplace are financial risk products introduced in the early 90's by Nestor Inc. with PRISM (http://www.nestor.com/), and HNC with Falcon (http://www.hnc.com/). Interestingly, they use radically different algorithms, each with key functional benefits for risk assessment. PRISM is based on a recursive adaptive model, and Falcon on a regularised MLP. They remain market leaders for credit card fraud detection [3].

Each system relies on much more than the neural network alone. In particular, the proprietary customer profiling pre-processors are the key to much of the generalisation power of the systems. At the user end, each system has the ability to feed explicit rules into a rule-based advisory system for transaction control, with the potential for direct blocking of credit card transactions, and GUIs are have also been crucial to the acceptance of this technology.

US Patents for intelligent systems

An indication of the commercial potential of neural networks among methods for adaptive computation can be gauged from a Boolean search of the US Patents Database (http://www.uspto.gov/patft/index.html), shown in Table 1.

Table 1
Results from a Boolean search of the US Patents Database over a 10 year period

Year	Artificial Neural Networks	Fuzzy Logic	Computational Intelligence	Genetic Algorithms	Adaptive Computation	Expert Systems
1987	-	3	-	1	4	2
1988	5	-	1	-	5	5
1989	7	7	3	-	17	17
1990	45	5	4	1	53	28
1991	58	13	3	1	71	19
1992	110	19	5	1	132	33
1993	136	55	12	2	201	30
1994	132	57	4	2	191	36
1995	182	65	9	4	254	29
1996	204	93	14	3	296	27

The term adaptive computation comprises the aggregate of patents represented in all of remaining columns except for expert systems, accounting for patents overlapping different columns. The table indicates the considerable growth in neural network and fuzzy logic methods, compared with expert systems based on predicate logic. Many of the expert system entries consist of high-level supervisory systems to validate or regulate assemblies of neural networks and signal processing modules. The power of these patents lies in the low-level processes, while their usefulness as a tool relies on the higher level control and interpretation provided by fuzzy logic processes, managed by the supervisory expert system. It is clear from the table how the rate of patent development using generic knowledge discovery tools is accelerating rapidly, and that the potential for new commercial developments is still in its infancy.

Expert systems patents continue to develop at a steady rate, and there is also a growing reliance on traditional methods of computational intelligence. The latter were taken to comprise natural language, parsing and constrained resource allocation, and will in future also include intelligent agents. Optimisation tools, typically using genetic algorithms, are also rapidly expanding, and this is likely to accelerate with the introduction of SWARM methods.

In conclusion, practical neural network systems have been commercially successful, even undergoing certification for use as medical devices. Neural networks have made an unusually rapid transition from laboratory into the marketplace, overtaking other better established technologies. They are mostly used under human supervision, or integrated with expert systems and fuzzy-logic systems. In later sections, stand-alone applications will also be described.

2.3 THE DESIGN LIFECYCLE

The lifecycle for computer systems with neural network components may be expressed in any of the standard ways. An appropriate representation based on that used by the FDA in their software guidance for reviewers and industry [4], is shown in fig. 3.

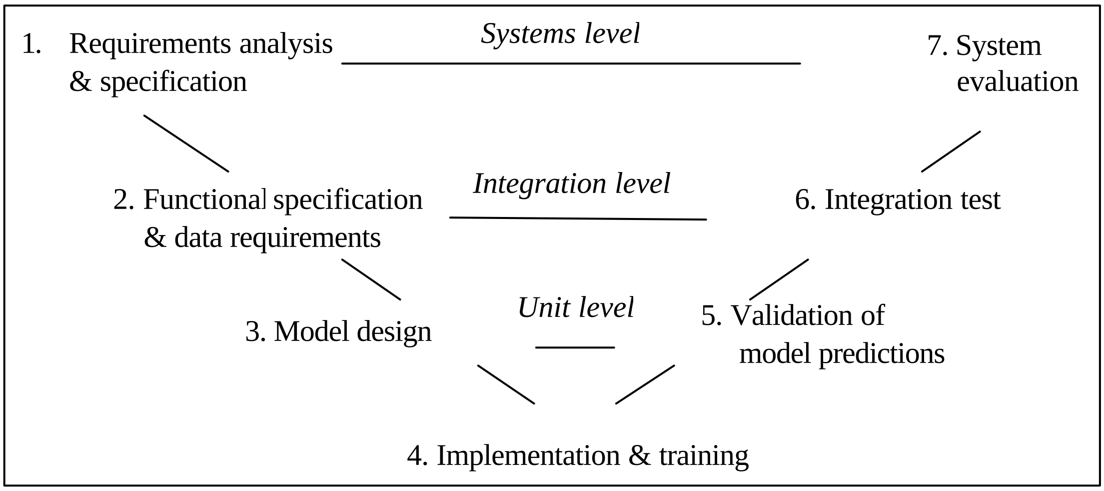

Figure 3
A generic software lifecycle model, with verification lines running horizontally.
A neural network is an empirical model, so the model design and data requirements take-over the rôle of knowledge representation and acquisition in symbolic reasoning.

A useful approach to good practice to promote confidence in advanced computer systems is 'to integrate assurance methods that lead to high confidence into standard systems development' [5]. This requires assurance methods to be applied at each level in the system development lifecycle.

Taking as an example the design of a non-linear process control module, at the unit level, controllers must be validated against established principles of stability margins and performance bounds. Extending the assurance principle to static pattern recognition, a performance assessment of the neural network implementation requires confidence estimates for the network predictions, possibly supplemented by novelty detection, which affords more specific protection against extrapolation from the design data. These requirements loosely reflect the concerns with completeness of the knowledge base and consistency of inference, familiar from prepositional logic.

It follows that some of the critical issues of particular relevance to neural networks are the need to focus in areas where significant data are available, difficulties with scaling systems from prototypes to routine use, and evaluation on real data. There is evidence that data collection, pattern representation and data integrity checking can take up a large proportion of the development time. A joint report by the Neural Computing Research Group at Aston University and Lloyd's Register (*http://www.ncrg.aston.ac.uk/cgi-bin/tr_avail.pl?trnumber=NCRG/97/009*) identified, among substantial issues for assessment of neural networks, the need to characterise data quality, relate this to defined safety margins, and eventually to trace neural network design features to the overall specification requirements at the system level.

At the integration and systems levels, the hazard analysis converges with the standard approaches. However, there are specific issues for systems with neural network components. To quote NASA [5]:

> 'Current software practices are human-centred activities. Such practices do not address very large or complex systems [to the extent that] novel software architectures using artificial intelligence and neural networks are stretching the behaviour understanding that can be achieved through state-of-the-art software validation and verification technologies.'

This points to the need to re-evaluate, for computational intelligence, how to impose discipline on design to minimise the introduction of additional complexity. With neural networks, at the unit level this requires parsimonious designs, and the integration level it involves reconciling the empirically derived optimal responses, with the structural models more naturally suited to represent expert knowledge. Given that 50% of errors are introduced at the requirements stage, a related issue is how to take a system-level view including human factors [5]. In the same report, the FDA calls for a set of expected best practices and complete hazards analysis of information-based safety-related systems.

The lifecycle is much the same as for decision systems involving statistical modules although for neural networks the technical aspects of verification at the systems, integration and unit level, are not yet established. There are also parallels with the corresponding stages in the design of knowledge-based systems, and indeed to any inference system with substantial non-linear components, whether using symbolic or distributed knowledge representations. Tracing through the blocks in fig. 3:

1. Expressing system requirements involves specifying against unwanted behaviour, in responses to unforeseen sequences of events. Many applications are now targeting environments that cannot be regarded as closed, and for which knowledge representations will necessarily be incomplete. Medical diagnosis is an early example of this.
2. Knowledge representations impact on generalisation ability, i.e. correct operation for future situations. In particular, human expertise is not always consistent and complete, and can be difficult to capture into an algorithmic representation.
3. It is interesting that there appears to be a convergence of knowledge-based, neural computing and statistical modelling approaches. The focus is on Bayesian models, where prior knowledge is used to define structural relationships, and statistical models are situated at the nodes, forming a multi-layered structure similar to a sparse version of that described in fig.1.
4. Assessing convergence is the network equivalent of achieving consistency. It can be established on the basis of statistical or computational learning criteria, to ensure appropriate coverage of the data without over-training.
5. There has recently been some controversy over the development of statistical models for medical diagnostics, emphasising the need for independent assessment by agents external to the original design process. This is partly the need to ascertain and automatically signal if the inference is extrapolating outside, rather than interpolating within, the knowledge base. This is a well know issue for neural networks, and while error bars go some way towards addressing novelty detection, it remains an acute problem when parsimony due to model selection causes data from none of the classes to map into highly predictive regions from one of the assumed classes.
6. Transparency of inferences is difficult for any complex system, and particularly so when knowledge is distributed. Nevertheless, in most current applications the optimised neural networks are sparse in the number of input variables used. Their small size allows their operation to be sufficiently traceable by direct inspection of the weights and hidden node activations in response to specific test patterns, to enable a verification against established domain expertise.

7. The crux of software regulation may be summarised as validation, verification and testing (VV&T) [4]. Any system may in principle be consistent and complete by design, yet contain knowledge that is incorrect. Therefore, while VV&T requires adherence to formal methodologies at each level of the design lifecycle, where non-linear inferences from real-world data are involved the emphasis appears to be shifting towards extensive trials with external data. This leads to a verification process that is performance-based, rather than founded on internal logic. A key element here is the need to define the required performance targets at the specification stage, rather than continually adjusting them on the basis of results obtained with network prototypes.

Progress in neural networks relating to the issues raised in each step of the lifecycle is discussed later in the report. In relating neural networks to other artificial intelligence methods, it is useful to place them also in the wider context of non-linear signal analysis, whose rôle they are sometimes designed to implement. The schematic in fig. 4 extends the discussion to a wider spectrum of models for information processing, on the basis that all of the methods listed involve selecting information for a parsimonious inference model. In this selection process, much information is lost, and the theoretical principles behind the system design and analysis are there to provide assurances about the accuracy and consistency of the predictions made in response to future queries, whether expected or novel, falling within or outside the system specification.

In most practical applications in the process industries, a continuum of models is used as a cascade, from conventional to novel engineering analysis, through core neural network systems, to supervisory knowledge-bases systems enforcing some form of self-validation.

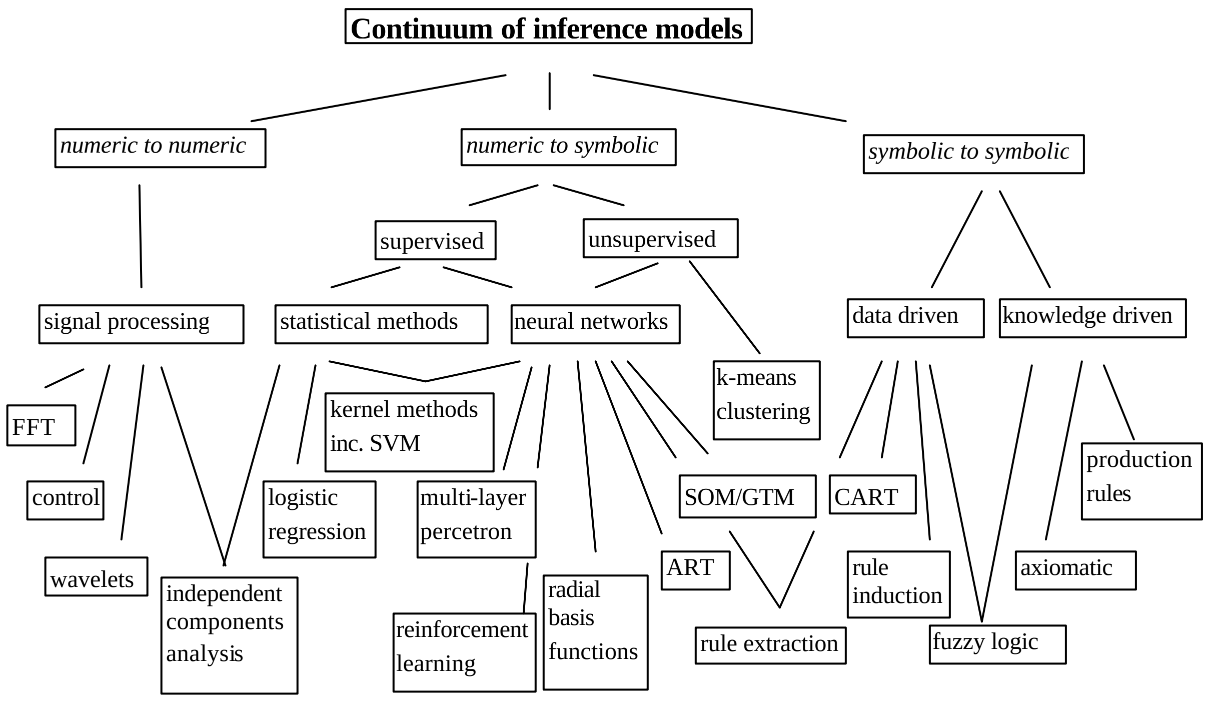

Figure 4
Global context of inference model. Some of the assignments are necessarily ambiguous and the list is not exhaustive, e.g. Bayesian models with latent variables straddle all three main headers.

3 USES OF NEURAL NETWORKS IN SAFETY-RELATED AREAS

3.1 POWER GENERATION AND TRANSMISSION

GNOCIS (Generic NOx Control Intelligent System) was developed by Power Technology for use as an on-line advisory or closed-loop supervisory system for NOx emissions (http://www.powertech.co.uk/enterprise/r&d_portfolio/gnocis_art.html). The purpose of the software is to adapt to long-term changes in the plant condition, enabling better optimisation of the operating mode of the plant. Trials were conducted at a 500 MWe unit at Kingsnorth power station, claiming to identify major annual efficiency savings by reducing carbon-in-ash, worth more than £100k per annum, while maintaining NOx emissions under prescribed limits. GNOCIS has now been applied to a range of boiler sizes, with several close-loop applications, with substantial efficiency gains.

In a separate project jointly funded by BCURA (British Coal Utilisation Research Association) and the Department of Trade and Industry, a hybrid neural network based controller for a 3.7 MW_{th} (i.e. MW thermal) chain-gate stoker-fired shell boiler at Her Majesty's Prison Garth, Leyland, in collaboration with James Proctor Ltd. This demonstrated 10% lower NOx emissions without sacrificing carbon-in-ash losses and with a 10% reduction in CO emissions, in steady state, with gains also when load-following. The code was implemented in Matlab, and expert knowledge had a key rôle for the integration of the neural network module into an efficient control loop structure (http://www.dti.gov.uk/ent/coal).

A related field with potential in power generation is soft sensing, meaning software to estimate specific process variables that cannot be directly measured, for instance pulverised fuel flow to the boiler. In control terms, soft sensing represents a non-linear extension of linear observers, which are typically Kalman filters. In addition to efficiency gains from tighter control this measurement would also provide fast indications of dangerous conditions, such as mill emptying by coal sticking to the feeder chains. Neural network based optimisation is recognised as a generally applicable technique with potential benefits for several continuously monitored loops in power stations.

Load demand forecasting is an early focus to demonstrate the potential of neural networks in practical problems from the power industry. There are at least two commercialised load demand forecasting systems, namely Nostradamus, from NewEnergy Associates (http://www.newenergyassoc.com/html/body_nfg_press_release.html) and ANNSTLF (Short-Term Load Forecaster, Version 3.0 http://www.epscweb.com/nonepsc_annstfl.html) marketed by EPRI. These software packages are used for daily and hourly electricity and gas demand forecasting. In the case of National Fuel Gas Distribution Corporation, in Buffalo, the forecast is for mode than 730,000 customers in Western New York and Northwest Pennsylvania. ANNSTLF claims to be implemented at over 30 electric utilities and to be used in real production environments, promising to cut operating and fuel costs by 5%. Since 1995 the Pennsylvania power & Light company has used ANNSTLF as its only forecasting tool for power sales. Reductions in the "power cushion" is projected to save the utility US$4 million in the next ten years.

Accuracy and economic value of neural networks for short-term electric demand forecasting, and for combustion optimisation with reduced NOx emissions, both featured at the 1998 America Power Conference, and the 1999 International Business Forecasting Conference.

Savings achieved by Duke Power Company in its inspection of reactor core control assemblies (RCCAs), from time reductions in the analysis of the 800 Mbytes of data resulting from a single core inspection, are estimated at $28,000 per inspection (http://www.nuc.

berkeley.edu/thyd/ne161/rtse/dukerods.html). This is because only 5% of the data contain relevant wear information. Together with the Knowledge Based Technology Applications Centre (KBTAC) of the Electric Power Research Institute (EPRI) Duke Power developed an Intelligent Data Reduction and Processing system (DEK-IDRP) that uses neural networks and rule based expert structures. The emphasis in the DEK-IDRP is on near-real-time advice with user-friendly GUIs, and an evaluation has claimed the detection of previously missed wear features. The operation of this system to analyse data from pressurised water reactors is expected to save 'US$361k in the next 5 years'. This application is typical of a staged approach to safety-related automation, where an advisory system is used simply to flag-up potentially important information, leaving the user to decide what information is actually important. This philosophy of operation is similar to that employed by Papnet, both systems having the rôle of helping human experts to sieve through a mountain of information, looking for a small number of tell-tale signs. This high sensitivity comes at a price of low specificity, which is resolved by the involvement of the user.

Nuclear reactor surveillance and diagnosis was the subject of a neural network benchmark study carried out in 1995 under the auspices of the OECD-NEA. In a separate project funded by the US Department of Energy, an artificial intelligence fault-diagnostic system for real-time detection of component failures was developed with data from Duane Arnold Nuclear Power Station. The OECD-NEA benchmark (http://www.nea.fr/html/science/rsd) was to predict the generated electrical power from a two-loop pressurised loop reactor at Borssele in the Netherlands. Some accuracy was reported except during condenser rinsing and stretch-out during shutdown operation. However, there was a recognition that a MLP trained with the data collected from plant does not adequately capture the process dynamics, especially when it targets one-step-ahead prediction of process output from the control variables. This is now well accepted, leading several groups towards using non-linear combinations of standard linear dynamic models, which can be identified locally with analogues of PRBS. In a follow-up report a year later, it was concluded that there are 'serious limitations to the effective utilisation of the MLP in practice', and that the benchmark results 'reveal the importance of establishing general guidelines for enhanced network training and effective utilisation'.

3.2 PROCESS INDUSTRIES

A manufacturing area where neural network control has been successfully applied for some time is steel rolling mills. Having developed prototype neural-network models for strip temperature and rolling force at the hot strip mill of Hoesch, in Dortmund, in 1993, Siemens has applied this technology at 40 rolling mills world-wide. Claimed efficiency gains are 30% better accuracy in rolling force modelling, and with prediction improvements leading to US$200k p.a. in material costs. The considerable business benefits demonstrated in routine use by a major manufacturing company, are indicative of the industrial push to explore new, non-linear, technologies. This application also illustrates two generally applicable exploitation areas for neural networks in process control. The first is to interpolate parameter settings more accurately than is possible with inheritance tables. This serves to close the gap between principled analytical models, and the effect of unmeasurable parameter disturbances, the network providing correction factors, in effect to calibrate the analytical model from batch to batch. The second growth area to automate systems of increasing complexity, perhaps comprising a number of interacting units, and often involving switching between different operation regimes. It is interesting to relate Siemens' experience that standard MLP-type feed-forward networks were not successful and purpose-built models for dynamic control had to be designed. These models are claimed to allow robust on-line adaptation, compensating for process drift even in multivariate processes.

Returning to the development lifecycle, Siemens indicates that neural networks always complement, and never replace, physical models. This statement implies that returning to basic principles is key to the interpretation of the non-linear models generated from the data, and hence critical to the safe, as well as efficient, process operation. Secondly, domain expertise, here in the form of linear and non-linear process control theory, is essential in the validation process. Thirdly, the data requirements are severe, taking thousand of strips, over several week's production time, to build a representative data set. All of these factors point to the issues that may be involved in verification of neural networks in safety related process control applications.

At Siemens, 'intelligent controllers' have also been applied successfully to coat thickness in hot dip galvanising lines. Furthermore, there are indications that these methods may be applied to the design of traditional proportional, integral and derivative (PID) controllers for plant with time delays, which would considerably broaden the scope for their application.

A generic third strand of application is in fault detection and identification (FDI). While frequently not a safety-related area, nevertheless it, too, has the potential for widespread use of non-linear dynamic models, as well as static and self-organising neural networks, serving various predictive functions from which departures from normal operation may be inferred. Despite some theoretical work in this area, visualisation of process variables remains fraught with difficulties because of the often severe dimensionality reduction that is required, which is exacerbated by the expedient dependence on heuristic methods. In contrast with soft sensing, which targets a specific process variable that is hidden from direct measurement, fault diagnosis often demands the simultaneous monitoring of a large number of variables, to detect departures from an ill-defined normality envelope. It is, in essence, a novelty detection problem.

3.3 TRANSPORT INDUSTRIES

Aircraft icing is a major hazard for which weather forecasters must advise pilots. The Experimental Forecast Facility at the Aviation Weather Centre in Kansas City, Missouri, is currently evaluating NNICE, a neural network-based icing intensity predictive forecast tool (http://www.awc-kc.noaa.gov/awc/nnice.html).

Also in the US, at NASA's Dryden Flight Research Centre, Edwards, a joint programme with Boeing is testing neural network damage recovery control systems for military and commercial aircraft (http://www.usc.edu/dept/engineering/TTC/NASA/newsarchives/apr_99control-software.html). The purpose of the research is to add a 'significant margin of safety' to fly-by-wire control, when the aircraft sustains major equipment or systems failure, ranging from the inability to use flaps to encountering extreme icing. Example aircraft where this approach can be applied are the Boeing 777, and the current test plane, a F-15 with canards and pitch/yaw vectoring nozzles.

At Long Beach airport inductive loops are used to identify aeroplanes at specific locations on the runways, using Loop Technology (LOT). The potential for use of low-cost surface sensors in avoiding incursion incidents relies on neural networks to classify loop induction signatures for accurate aircraft type identification.

In the UK, vibration analysis monitoring in jet engines is the focus of a research project involving Rolls-Royce and the Department of Engineering at Oxford University. This produced a diagnostic system, Quince (http://www.eng.ox.ac.uk/World/Research/Summary/B-Neural.html), which combines the outputs from neural networks with template matching as well as statistical and signal processing methods, processing them with a small set of rules. The software is designed for the pass-off tests of jet engines, has a tracking facility to suggest the most likely fault, and centres on the use of novelty detection to identify unusual vibration signatures. According to the web site, Quince is now being licensed to Rolls-Royce under the terms of a Licensing Agreement signed in May 1998.

A second condition monitoring application between Rolls-Royce and Oxford University involves predicting a thermocouple reading of the exhaust gas temperature in aero-derivative gas turbines with a power output of 3-50 MW. High prediction errors are indicative of developing faults, and it is claimed on the web site that the model is capable of identifying real faults several hours before it is detected by the control-system logic which shuts-down the engine.

A third collaborative application listed on the web site is to perform comprehensive whole-engine data analysis and interpretation, with attached confidence levels, by fusing diverse sensor readings (performance parameters, vibration spectra and oil debris information) to produce 'reliable indications of departures from normality'. The aim is real-time in-flight monitoring for the new Trent 900 Rolls-Royce engine. Technically the project combines standard observers, i.e. Kalman filters, with more advanced signal processing techniques and neural networks, as well as other elements of computational intelligence.

European collaborative projects in Framework Programmes 4 and 5 have implemented neural network control demonstrators, ranging from engine management models to physical speed control, and involve leading car manufacturers. The key feature of these control systems is the combination of detailed engineering expertise with non-linear interpolation by neural network architectures. In specialist applications such as real-time control of complex dynamical systems, sometimes requiring rapid adaptation to operational environments, and always susceptible to re-tuning for production models and over time during routine maintenance, it is likely that dynamic models designed from first-principles will be at the core of controller design. An important consequence of this approach is that closed-loop stability can be theoretically established, even when switching between different operating regimes, themselves operating under the guidance of supervisory neural-, fuzzy- or rule-based systems. This is a good illustration of the smooth coupling of a range of inferencing methodologies, that is key to many important practical applications where novel solutions are needed. In process control generally, stability can now, in principle, be assured for complex combinations of very different models, operating in the real-world with a wide range of disturbances and uncertainties. Examples of good practice may be found in projects NACT (http://www.mech.gla.ac.uk/~nact/nact.html), FAMIMO (http://iridia.ulb.ac.be/~famimo/) and H2C (http://www.control.lth.se/H2C/).

3.4 BUILDING SERVICES

Siemens currently markets the FP-11 intelligent fire detector. This uses 'FirePrint' technology, which is based on fingerprints, that is to say time traces from different types of sensors. These were acquired from fire tests carried out over many years, resulting in extremely high specificity, triggering one-thirtieth as many false alarms as conventional detectors (http://www.cerbpyro.com/deintel.htm). The developing company, Cerberus, a division of Siemens Building Technologies, are sufficiently confident about this product, also called AlgoRex, that they have offered a refund for the costs of any unnecessary fire department visits triggered by this alarm. The detector's appearance is similar to an of-the-shelf fire detector, mounted onto a bulky base. It offers three user-selected options for the system logic, using an optical sensor to detect smoke density, a heat sensor for temperature, or the neural network which combines both measurements. In reality, the network is based on a digital implementation of fuzzy logic, with rules discovered by the neural network but validated by human experts. According to the manufacturers, 'the rules used in the fuzzy logic system are the result of decades of know-how gathered by experts at Cerberus from approximately three million AlgoRex fire detection systems in operation world-wide'.

A separate product under development is the WaveRex flame detector, which operates in high ceiling sites where the amount of smoke would not necessarily trigger a detector. This combines a fuzzy logic system with a wavelet filter bank, to distinguish profiles of real fires from those of reflected sunlight. Flicker measurements are made at 4.3 μm (CO_2), 5 and 6 μm (black-body radiation) and 0.8 μm (visible range). The product sheet claims that this is the first time that wavelet analysis is used in a mass produced article (http://www.siemens.com/Ful/en/zeitschrift/archiv/Heftl_99/artikel03/index.html).

A further generic area where there is much demand for new products is smart buildings, which extends from semi-automation of household appliances to remote care for convalescing or elderly patients. At Siemens, active vision is being developed, through cameras with built-in computers that capture event profiles for automatic detection of the event even under variable lighting conditions. Neural networks are likely to be at the centre of this.

3.5 CONSUMER PRODUCTS

The Sharp LogiCook is the first microwave oven to be based on neural network technology, developed with the Oxford University (http://www.eng.ox.ac.uk/World/Research/Summary/B-Neural.html), and is also claimed to be Sharp's first product developed outside Japan. The user needs to specify whether the cooking material is food or drink, to distinguish between cooking and heating. Its temperature is then uniformly raised to 75 degrees Celsius, and the optimum cooking time is obtained from an analysis of the proportional, integral and derivative humidity profiles, carried out by a neural network operating on a Hitachi micro-controller. The LogiCook can deal with frozen, pre-heated and different sized portions of the same food. It is also capable of detecting the dangerous condition of liquids, superheating, and can issue stir commands as necessary (http://www.scit.wlv.ac.uk/~cm1822/acnl7.htm).

3.6 HEALTHCARE

The earliest and currently most widely used neural network based system in healthcare is Papnet (see 2.2). There is widespread credibility that this testing support software improves detection rates for cervical cancer from Papnicolau stained smear slides, but the debate about cost-effectiveness still goes on. However, it terms of regulatory requirements this software has lived-up to the claims made.

A good example of a hybrid decision support system in healthcare is GLADYS (GLAsgow system for the diagnosis of DYSpepsia http://students/dcs.gla.ac.uk/students/lamkc/CPI.html), developed by the Glasgow Southern General Hospital with support from the University of Glasgow's Department of Public Health. This is a Bayesian model for the diagnosis of several conditions relating to dyspepsia, built using a combination of knowledge acquisition from clinicians, and statistical representations to encode that knowledge in a structural form that can be updated numerically, and used to process uncertain knowledge in a consistent manner. An extension of this approach with Markov Chain Monte Carlo integrals to ascertain in more detail the precise shape of the probability distributions used in the inference process, would today be directly at the interface between statistics and neural networks. Yet the system was originally introduced in 1990 as an expert system, on account of the key rôle of knowledge extraction in defining the model structure, and verifying Bayesian relationships constructed by the system [6]. At the time of publication, GLADYS was used in several hospitals across Europe. During 1989-90 it was the subject of a prospective trial involving 8 UK centres, based on a design database of 800 patients. The claimed benefits of the system included i) success in eliciting information from patients, through a direct interface that does not involve a clinician, and ii) a reduction in referrals by GPs to hospitals conservatively estimated at 30%, 'with no drop in the standard of care'. Most of the development time was concerned with implementation programming, system refinement, data enhancement, and testing, totalling 21 man-months from the total quoted of 27 man-months. The remaining 6 man-months were taken-up with interviewing medical experts, as well as statistical analysis and design. An important design aspect of the system is that it does not purely provide a diagnostic prediction, for up to 27 disease classes. It also identifies symptoms that indicate that the disease is present, or provide evidence that it is absent, and which symptoms do not contribute to the disease, besides handling the possibility of several diseases being present. A recognised risk of the system was the potential for over-reliance on what was intended as an advisory, rather than executive, system, and this is more prominent when used by non-experts, where the major benefit in reducing unnecessary hospital referrals may be.

Questar is a sleep analysis package, developed initially by Engineering Department at Oxford University (http://www.eng.ox.ac.uk/World/Research/Summary/B-Neural.html), initially marketed by Oxford Instruments, and now taken-over by a recent spin-out company from Oxford University, Oxford BioSignals. According to the web site, the software name stands for Quantification of EEG and Sleep Technologies Analysis and Review. It started-off as a EPSRC programme, developed with a DTI-LINK funding, was launched in September, 1995, and sold from April 1996. It was awarded a British Computer Society medal in 1996, and gained FDA approval in 1997.

The purpose of the software is to automate sleep staging into awake, rapid eye movement (REM) or light sleep, and deep sleep, as accurately as an expert user, but on a continuous scale and with a much faster sampling rate of 1Hz. It does this by combining three electrical measurements, Electro-encephalogram (EEG), electro-oculogram (EOG) and elector-myogram (EMG), which measure mental activity, eye movement and muscular activity, respectively.

A second healthcare project at Oxford University is the development of a software monitor for intensive care patients. This area of medicine is fraught with controversy, since demand for intensive care beds is very variable, yet the costs involved are very high. As a result, a decision may have to be made whether or not to admit a critical patient into intensive care, sometimes with the aid of statistical advisory systems, the most commonly utilised of which is Apache II, whose development was led by Glasgow University. Further software is also used in several hospitals across Europe to aid in the management of critically ill patients. Interestingly, a Bayesian model of clinical data has been used to test the hypothesis that Cerebral Partial Pressure does indicate the presence of sub-clinical damage by trending during the first 24 hours following admission. This indicates that careful monitoring of this highly invasive measurement can improve the management of patients who ‘talk and die’, i.e. score highly at admission e.g. with Apache, but subsequently show poor prognosis. This is an example of the use of rigorous statistical methods applied to neural networks, for knowledge discovery in medicine.

In its original form the ‘Software Monitor’ processes five standard physiological measurements, namely EEG, systolic blood pressure and oxygen saturation, breathing rate and temperature. It uses these standard non-invasive measurements to alarm for novelty. It is an example of a data based system where the available signals define one state, normality, which is not the event of interest. The difficulty lies in accurately and robustly identifying deviations from the multivariate region enveloping normality, since when simultaneously monitoring several variables, the population density in the input space is very low. Moreover, the process must accommodate frequent updating of the software in response to the changing ’normal’ state of the patient as they gradually recover consciousness, while excluding artefacts arising, for instance, from sensor displacement due to patient movement.

Another patient monitoring system previously tested at St. James' Hospital, Leeds [2], with a continuous run over 3 months, also identified the involvement by the user, in this case nursing staff, as a significant element in the system operation and one that requires careful analysis in its own right.

Two further topical areas are the control of anaesthesia and monitoring of awareness during surgery. A fuzzy logic controller for anaesthesia by the Department of Engineering at Sheffield University is undergoing clinical trials at the Royal Hallamshire Hospital, Sheffield. Depth of anaesthesia is monitored by fusing cardiovascular measurements with parameters derived form the auditory evoked response, to control continuous infusion of the anaesthetic agent Propofol [7].

Auditory evoked potentials are EEG traces that are time correlated with auditory signals used to stimulate the brain of unconscious patients. There is currently much interest in using them for real-time monitoring of depth of anaesthesia. This was developed by Pacific Northwest Laboratories (http://www.emsl.pnl.gov/), in collaboration with Electro-Cap. While a full clinical evaluation is still very much in its initial stages [2], this project is indicative of healthcare trends.

In a different line of development, a recent project led by Lund University reached the stage of clinical trials for a web advisory system for automated intepretation of myocardial perfusion images, returning diagnostic advice within seconds (http://www.weaidu.com /software/index.html). The group is experienced in the design of signal processing and neural networks for medical advisory systems for large-scale use, having previously tested an acute myocardial infarction (AMI) detection system on a data base of 1,120 ECGs from patients with AMI and 10,452 control ECGs. In that application, the neural network system was found to be more sensitive and have a higher discrimination accuracy than benchmark ECG software, or expert cardiologists. Establishing the future use of such systems needs careful appraisal in a much wider context of care costs patient management priorities at particular hospitals, but the projects nevertheless highlight potential benefits is raising baseline standards through intelligent data access using pattern recognition methods.

3.7 ENVIRONMENTAL MONITORING AND CONTROL

An example of this is an application from Unilever's Environmental Safety Laboratory, modelling corrosivity of organic acids (http://www.spss.com/spssatwork/stories /unilever.htm). Claimed benefits include a reduced need for animal experimentation. Another example is the Ministry of Agriculture Fisheries and Food to predict toxicity of food additives, in order to flag substances that need further investigation (-idem-/stories/maff.htm). A current practical system from Odin Corp detects engine misfires at engine speeds of up to 10,000 rpm in real-time. Misfires are believed to be a leading cause of pollution, to the extent that the California Air Resources Board has mandated that by 1994 all new cars must detect misfire in real-time. The Odin system requires 3 Kbytes of software running on a Motorola 68030 microprocessor (http://www.dacs.dtic.mil/techs/neural/neural11.html).

There is a major effort to develop commercial electronic noses. Part of the impetus for this is the high profile of neural networks, since it offers a tool to unscramble the complex signatures produced by large arrays of tuned chemical sensors. In a technical brief from Pacific Northwest National Laboratory (http://www.emsl.pnl.gov:2080/proj/neuron/briefs/nose.html), three different approaches were identified, relying on conventional chemometrics, artificial neural networks, and advanced biological models of the olfactory system. The range of opportunities is vast and profitable, and some of it safety sensitive. Current applications of electronic noses, for instance using the commercial system zNose (http://www.estcal.com /Technical.html) include drug and explosive detection, environmental monitoring for pollutants and toxic chemicals, and nerve gas detectors. It is noteworthy that this system produces ‘visual olfactory images’ that are interpreted by the user, not by the system alone. These polar plots of odour intensity can allow visual identification of trace elements of almost anything from bacteria to gasoline, and the ‘olfaction’ label results from the use of chemical sensors rather than gas chromatography or IR spectroscopy, or other industry standard measurements for the same substances. The zNose claims to have been verified by the US Environmental Protection Agency’s (EPA) Environmental Technology Verification programme (ETV). Other electronic nose systems do involve on neural network discriminators (http://www.cogs.susx.ac.uk/lab/nlp/gazdar/tech/atc/1998/web/sloss), but it is not clear whether these systems have been certified for safety-related use. A recent conference (http://www.knowledgefoundation.com/events/8150902.htm) raises the question of the standards needed for the validation of eNose measurements, since the sensors involved, a combination of electronics and ceramics technologies, are prone to large drifts and making it difficult to carry out essential tasks such as accurate calibration.

Vusion, a new company in the Austin Technology Incubator at the University of Texas, is preparing to market an electronic tongue. The potential quality monitoring applications for the device range from pharmaceutical testing, for instance for cholesterol, to quantifying the freshness of perishable foods. (http://www.businessplans.org/Vusion/Vusion00.html).

4 POTENTIAL SAFETY GAINS FROM NEURAL NETWORKS

Although applications of neural networks range across a large number of industries, in the vast majority of current applications their rôle fits into two generic classes, each with further specialisations.

4.1 STATIC PATTERN RECOGNITION

This is the original motivation behind the exploitation of neuromorphic systems, but developments have forked into two distinct routes. The vast majority of image processing, sensor fusion and diagnostic systems use static neural networks whose function is entirely statistical by nature, and whose rôle is that of distributed, non-linear, associative memory models. When they are regarded in this manner, their benefits, associated hazards, and regulatory requirements could be expected to align with those for linear multivariate statistical methods, but the theory underpinning robustness is much less developed for non-linear than for linear statistics. Potential safety gains are to be derived form automation of diagnostic and monitoring systems, as well as greater accuracy, usually expressed in terms of very specific performance requirements for true detection rate (sensitivity) and false alarm rate (specificity).

A second strand of current developments follow-on from our nascent understanding of neuromorphic systems, which is the case of mammalian sensory networks has become very detailed. For instance, the physiology of the retina is sufficiently well understood at the level of axonal signals, to enable electrical replicas to be constructed, whose dynamic range and stability under variable light conditions are extremely good. Beyond this, at a cortical level, simple and complex cortical cells have been directly mapped for their frequency response characteristics, and can be emulated using banks of wavelet filters combined into quadrature pairs. A commercial example of this is an iris classification system for access control (http://www.cl.cam.ac.uk/~jgd1000/). With wavelet-based whole face recognition it is possible even to characterise facial expression. Auditory and olfactory systems have also been modelled in detail, and their performance potential is huge. Potential safety benefits arise from the ability to place near-human performance in locations or environments where it is either impossible, or too costly, to employ people. As the hardware implementation of 'bionics' develops further, there is clear potential for prosthetic implants, as well as the design of circuits that exceed certain aspects of human sensory performance, notably speed of response. An example is the development of artificial retinal implants (http://rleweb.mit.edu/retina/).

Further practical benefits from emulators of cortical circuits are sparse signal representation and excellent signal-to-noise differentiation. It is not difficult to conceive of safety-related applications reliant on 'bio-sensors' coupled with 'intelligent' supervisory systems, for instance for fire or smoke detection in remote environments, in the presence of variable conditions such as exposure to inclement weather, requiring neuromorphic processing to maximise the contrast between the signal sought and a background of continually varying disturbances. Yet another potential gain is in human-computer interaction, where hand-written and verbal natural language interfaces are already being developed.

4.2 DYNAMIC CONTROL AND MONITORING

Immediate performance benefits can be gained where control with complex switching requirements is optimised by replacing parameter allocation look-up tables with smoothly interpolating non-linear models. In many cases this require identification and control of dynamical systems, a very difficult task where non-linearities are involved. Recent developments have resulted in theoretical frameworks to demonstrate stability for non-linear control, and these have arisen from combined developments in control theory and in numerical analysis. It is a telling development that traditional control expertise is even more necessary for the safe deployment of neural network controllers then it was for traditional three-term controllers.

The need for this expertise is apparent from a constructive approach to the development lifecycle. At the unit level, confidence should be built-in through design with stability and performance assurances. At the integration level, the availability of advanced controllers may enable, for instance the replacement of hydraulic actuators with electrical ones that are easier to build back-up systems for. At the system level, there are implications arising from the additional flexibility afforded by, for instance, drive-by-wire, where braking and suspension control can be inter-linked to correct lateral stability under heavy braking, to give but one obvious example. In another example, taken from the preceding section, fly-by-wire control may be automatically reconfigured to compensate for severe system failure. This applies equally well to commercial, as to military aircraft. Ultimately, safety benefits will be gained from tighter control of complex switching systems which exceed human performance in emergency situations.

Further engineering gains are likely to result from developments in intelligent sensors, for which demand from industry is substantial. Safety benefits will arise from access to estimates of hidden state variables. An example of this would be estimating the amount of pulverised fuel in a coal mill, when wet coal can become trapped in the rolling chains of a coal feeder. This results in fast emptying of the mill by the primary air flow, which is there to dry the coal, filter fine particles in a cyclone, and carry the coal directly to the burners. The measurement difficulty is that any probe to measure pulverised fuel flow quickly clogs-up, yet this measurement is important both to regulate the fuel efficiently, and for safety, since an overly rich air-to-fuel ratio is spontaneously combustible at the temperatures that these mills reach when running empty. In effect what is needed is a sophisticated non-linear observer capable to estimating the coal content of the mill for air temperature and differential flow measurements. This is just one example from many in the process industries where direct measurement of variables key to performance and to safety are either too expensive to measure accurately, or even impossible to measure fast enough for closed-loop control, for instance pH values in sulphonation loop reactors.

Plant monitoring is altogether a different type of system, where the aim is to compare actual plant operation with nominal operation given the current control actions. Any deviation is indicative of plant malfunction, which can trigger a diagnostic search for the most likely fault, in real-time. The benefit here is in earlier warning of incipient faults, as well as the potential for providing the operator with a ranked list of possible hazards, replacing a plethora of consequential alarm calls. The nuclear sector is a potential area for advisory systems of this nature, but the potential is much wider that this. The difference between using neural networks rather than purely knowledge-based systems for this application lies in the ability to integrate complex signals at a low level, yielding more complex advice than may be possible by representing the activity of individual sensors directly in symbolic form.

5. HAZARDS ASSOCIATED WITH NEURAL NETWORKS

5.1 STATUS OF NEURAL NETWORKS IN SAFETY-RELATED AREAS

The current safety-related areas where neural networks are in routine use are:

1. advisory systems for healthcare;
2. consumer products;
3. short-term load forecasting of electrical power and gas;
4. industrial process control and monitoring; and,
5. fire alarms.

In each case the status enjoyed by the computer systems is determined by two main factors, namely:

1. confidence in the process models used to design, validate and test the neural networks;
2. perceived benefits experienced during routine use, particularly in relation to adverse events.

There is a tacit understanding that reputable companies sell only reputable products, whatever technology they may contain. In the case of consumer products, labelling a product as containing an 'intelligent system' is sought after as evidence of advanced technology, implying better, and safer, performance.

Where the users are themselves domain experts, as is the case in healthcare, and in the utilities and manufacturing industries, routine use of novel technology is entirely rooted in their performance against existing benchmarks. While this may be justified when large companies have the expertise in neural networks to adequately verify them against domain knowledge, often writing purpose-built algorithms, a potential limiting factor is that many companies cannot afford the expertise to develop neural networks in-house. They rely on commercial systems that may not have the necessary regularisation procedures in place to prevent over-optimistic inferences being made, and whose results may be treated naïvely.

In addition to lack of transparency and propensity to overfit the data unless the network carefully controlled, there are other well-known concerns regarding neural network system components, which are summarised below.

5.2 NOVELTY DETECTION

Novelty detection simply maps onto neural networks the difficulties with detecting untypical plant states. This tests the completeness of the artificial intelligence model, since it is difficult to argue that any formal model is formally complete, when dealing with real-world data. In this respect, at least with statistical tests the completeness of the model can be assured for the design data, and the novelty of future data can be tested with reference to those same design data, resulting in a well-posed problem amenable to theoretically satisfactory solutions.

A related difficulty not yet fully resolved, is the extent to which the available data are representative of the range of expected behaviour from the plant. In a particular application, the impact of data quantity on the performance of neural network freeway incident detection models was the subject of a study in Australia (http://www.its.usyd.edu.au/). This is indicative of the concern that exits around data dependence of network operation, and the current lack of answers, despite some theoretical advances that are summarised in section 6.1.

5.3 EVALUATION AND TESTING

Empirical models necessarily rely on the data used to design them, which typically separate subsets for training, out-of-sample validation, and prospective testing. It follows that demonstrating performance claims will also rely on the collection of further empirical data. This is true for as much for statistics as for neural networks. However, when error models are not well established, which is currently the case for non-linear inference models including neural nets, then the demands of performance testing rapidly escalate. The evaluation costs are commensurably high, potentially dwarfing the development costs for the system.

Another important aspect of empirical modelling is the requirement for data integrity. There is anecdotal evidence that data artefacts may not the exception, but the rule. An example is where observations are entered manually into diagnostic devices, since the values read may be arbitrarily approximated by the reader. An example from the health sector is blood pressure readings, where a data set was found to contain only even values up to around the 100 mark, with odd as well as even above that figure. Apparently, it turned out that the dial is only marked in even steps, but nurses had been instructed to check high readings by averaging two measurements. In practice, a good health test for a diagnostic system is when it finds unsuspected artefacts in the data.

Other serious difficulties arise from mis-labelling, as well as missing inputs. As with expert knowledge, so a gold standard for class assignment may not be easy to specify unequivocally.

5.4 MANAGING THE LIFECYCLE

A major potential hazard for intelligent systems, with or without neural network components, is what happens after they are in routine use, that is to say safely managing their developmental evolution and maintenance.

It is well established that the key to successful transfer of neural networks to the marketplace is often successful integration with routine practice, rather than optimisation for the idealised environments where much of the current development effort takes place. This is nothing new, of course but, as a principle, it does have implications for the deployment of new technologies.

Two examples of this are given here, illustrated with reference to control of dynamical systems. The first is the adaptation of neural network controllers from simulated data, however life-like, to real physical systems. It is to be expected that fine-tuning will be required to address unexpected limit cycles, or other control performance inadequacies that may arise on plant. If the controller is a completely opaque black box, then the only way to carry out this necessary re-tuning is by wholesale re-training of the controller. Identifying the data requirements to correct an observed fault, even when the desired effect is readily understood, would be virtually impossible. Therefore it is necessary for the non-linear controller to be implemented as far as is possible, with traditional linear control blocks, which can be readily adapted to correct the performance in particular operating regimes and at particular frequencies. Without a transparent implementation defined at the functional level of systems design, the controller would be of limited use on any plant other than that used specifically in the design process.

Even when a black-box controller is designed for a particular large item of plant, and it is shown to deliver tangible cost benefits through performance improvements, it is not uncommon to find in subsequent visits to the plant, that it has been switched out of the loop. This situation will arise as plant wear and variations in different batches of raw materials, require the controller to be slightly adapted in order to match the benchmark controller, or even to remain stable. It is generally the case, that maintenance of neural network systems cannot be carried out by users not involved directly in the prototype development. At best, this can negate the benefit from substantial investment in this new technology, and at worse it could leave plant operating under unsafe conditions.

5.5 SYSTEM INTEGRATION

A potential hazard with any advisory system arises when they perform with high accuracy in decision support. The two most commonly quoted dangers are :

1. Over-reliance, where even specialist users become complacent about the accuracy of the advisory system, reducing their critical awareness of the machine's inherent limitations; and,
2. When non-specialist users rely on the system's performance to act beyond their normal scope of competence. An example of the latter would be if decision support systems were to be used for triaging in emergency wards, in areas where a specialist doctor is not immediately available. Recent studies about the epidemiology of medical error (http://www.bmj.com /content/vol320/issue7237/) have shown that decisions by non-specialists under severe time constraints can be identified as a source of mistakes. In the US, this annual toll is said to exceed the combined number of deaths and injuries from motor and air crashes, suicides, falls, poisonings and drownings.

It is not difficult to foresee analogues of this situation developing in industrial safety-related applications, directly arising from reliance on mechanistic intelligent systems.

Some of these dangers are recognised in the key recommendations of a European funded project called 'Towards Evaluation and Certification of Telematics Services for Health', TEAC-Health [8]. While primarily concerned with the standards of web-site information and telecare, the project also looked into software for clinical care. The recommendations also recognise the legal issues arising from conflicts between European and national interpretations of what constitutes a 'device', for regulatory purposes. In particular, software that transforms standard desk-top computers into decision tools is a contentious issue.

Nevertheless, in healthcare the European Directive for Medical Devices (93/42/EEC) requires all devices 'placed on the market' to satisfy specified 'essential requirements'. The manufacturer (i.e. the person or organisation responsible for the device) also has to go through the relevant 'conformity assessment procedure(s)'. The choice of procedures depends on the class of the device according to a set of classification rules. Most of these procedures involve a 'notified body' which is an independent certification body chosen by the manufacturer from a list of bodies designated by the competent authorities of the EU countries. The Directive applies both to software embedded into a device, and also to some free-standing software (for example if it converts a general-purpose personal computer into a diagnostic system affecting patient care). This will put pressure for a clarification of the legal position of non-medical safety-related software across Europe.

6 SAFETY ARGUMENTS FOR NEURAL NETWORKS

6.1 PATTERN RECOGNITION

There are theoretical foundations for design and training of the neural networks most commonly used in pattern recognition. Typically, this involves a Bayesian framework to estimate regularisation parameters, and to predict the width of the posterior distributions generated by the model - see for example the tutorial in [2]. A more demanding procedure is to carry out full Bayesian integration by importance sampling over the full space of parameters and hyper-parameters, with Monte Carlo methods. These methods are becoming increasingly efficient, and combined with rapid increases in low cost computer power, have made possible detailed analysis of complex probabilistic models. They form two levels of accepted good practice for individual neural network modules in safety-related applications, both centred on statistically founded estimates of inherent error in the model predictions. In decision support, this uncertainty is used to modulate the predicted class membership probability, towards the guessing line.

At the integration level, there are two alternative methodologies that have been used to promote confidence in decision support systems. One is novelty detection, which normally involves modelling the input density distributions, to signal whether new data are very unlikely to belong to the same distribution as the design data. A more recent refinement of this approach uses Extreme Value Theory to better quantify the probability estimates made at the tails of the density distributions. Either way, the emphasis in explicit novelty detection is to focus, not on whether the predicted decision is right, since that is part of performance testing, but on knowing whether to issue a prediction at all. The second approach is a variant of the established practice for safety-related hardware systems, of improving reliability through redundancy. This involves ensembles of neural networks, each fitted to independently drawn data samples. In an effort to reap profit from adversity, the very fickleness arising from convergence to local minima, for which neural networks are often criticised, can be shown to result in independent errors even when fitted to the same data, but from different random initial conditions. In other words, a range of theoretical results enables the constructive analysis of combinations of non-linear inference models, whose predictions are more reliable than from any of the constituent models.

Performance claims then need to be verified using statistical indices that are derived from extensive trials. A commonly used figure of merit is the area under the Receiver Operator Characteristic, and the FDA have explored bootstrapped methodologies to parameterise its dependence of the number of examples used for training, tuning and external testing of the neural network. Insofar as performance trials go, this approach is certainly comparable with the standard tests for linear statistical methods, and provides at least an educated guess at the sample sizes required to substantiate performance claims. This is the neural network equivalent of power calculations, that are so central to the design of safety-related tests, e.g. randomised case-control trials.

Ultimately, the preferred way to establish confidence in an empirical model of any kind, is to map it onto a structural model of domain expertise. This need for transparency motivates rule extraction from neural networks, to support symbolic reasoning. However, there are no *de facto* standards to determining explicit knowledge from that distributed in the network parameters. A common baseline approach is to carry out a sensitivity analysis of the network response, by differentiating its prediction with respect to variations in input values. In decision support, the derivative is best applied the logarithm of the odds-ratio, which is the probability of within-class to out-of-class membership. A more sophisticated approach is to use the hyper-parameters in a Bayesian regularisation framework to rank the input variables by relevance to the network's response, which is termed automatic relevance determination (ARD).

With neuromorphic engineering systems, verification is analogous to training of human operators, and may need to be carried out using standard hazards analysis for the electronic implementation, combined with performance tests to verify system functionality. The design lifecycle in fig. 3 still applies, but the functional specification and testing are now by reference to explicit computational or neuro-physiological models, therefore the hazards analysis will come closer to that to the standard systems approaches.

6.2 DYNAMIC MODELLING AND CONTROL

After a prolonged search for truly distributed non-linear dynamic models, it has become apparent that there are significant difficulties with sampling the state-space for accurate identification of transient dynamics in non-linear plants, and to theoretically underpin stability and performance bounds for black-box neural networks. Provable black-box neural networks are not yet feasible, unlike alternative practical representations based on combinations of locally linear dynamical models, such as the typical components used in traditional digital control.

When combining explicit linear models, neural networks provide a theoretical framework to demonstrate stability and evaluate stability margins, which is lacking in widely used heuristic approaches to non-linear control. While very successful in prototype applications, some of which have involved physical demonstrators, the 'feel' of these controllers still needs to be improved during transient responses, when switching between operational regimes. Further developments may require more radical design philosophies, based on robust control of families of functions, rather than optimising for a single plant. It is clear that neural network control is an area of active research, and significant results are continually being achieved.

An orthogonal development to dynamic identification and control with neural networks, is to use them for advisory purposes, either as supervisory systems, or off-line with advice about tuning of conventional controllers. For example, Siemens has issued a 'neural control cook-book' that includes the tuning of three-term controllers for plants with substantial time delays. These controllers clearly fall within the remit of traditional verification methods, since the stability margins and response of the resulting controllers may be theoretically verified against plant models. Difficulties will arise where the controller is automatically generated without the plant dynamics, and time delay in particular, being explicitly identified. Similar concerns were expressed in the response to the questionnaire to neural network practitioners appended to this report, about adapting neural networks on-line. Given that the adaptation, or 'learning' ability is perceived by many as the main selling-point of this technology, this a concern that must be taken seriously in safety-related uses of the technology.

6.3 IS CURRENT PRACTICE ACCEPTABLE FROM A SAFETY VIEWPOINT ?

The acceptability of current practice in neural network design reflects the main themes explored in this report, namely:

1. ensuring generalisation performance by avoiding over-complexity in the network architecture,
2. defending against blind extrapolation, and
3. evaluating not just their predictive performance but also their interpretation *viz.* explicit reasoning models, whether they are analytic or symbolic in nature.

These themes map onto standard certification requirements. The doctrine of 'substantially equivalent systems' would seem to require neural networks to be tested against multivariate linear statistical models, to justify the need for additional complexity. This can in principle be established by thorough benchmarking, which is the practice for the safety-related applications reviewed here.

The other standard doctrine is that the user must act as a 'learned intermediary'. This currently relies on a detailed analysis of the neural network function in response to a wide range of inputs during validation tests, to relate the empirical model to prior expert knowledge.

The current theoretical design frameworks for static neural networks carry out novelty detection implicitly, by exploding the predicted error bars if new data stray outside of the support of the design data. There are limitations to this approach, since the error bars depend on Taylor expansions and will saturate rapidly outside of the envelope of the design data. Whether or not further checks for novelty are need, remains an open question for this technology, and for computational intelligence generally.

Moreover, commercial software lags behind most of these developments, not least because they are fast removing accessibility of this technology to non-experts, who form a large component of the market for software development platforms.

It is a telling remark that the successful early pattern recognition products relied on good engineering design, that is to say careful selection of data representations, rather than statistical sophistication. It could be said that they have proved to be more accurate in routine use than might be expected given the simplicity of the final design.

In the control of dynamical systems, engineering expertise is always at the centre of successful developments. Current practice here is certainly fully consistent with that used for linear controllers. Taking into account the plethora of gain scheduling and other heuristics embedded into many industrial controllers developed in-house, I would personally feel more comfortable with the level of rigorous analysis that I have seen in the development of neural networks for the same purpose.

It is safe to conclude that behind each successful neural network product, there lay good engineering practice.

6.4 IS THERE A NEED FOR FURTHER GUIDANCE ABOUT BEST PRACTICE ?

The preceding section indicates that *de facto* industry standards for neural computing have become available, especially for static pattern recognition. Therefore, it is feasible to put together expert groups to express generally accepted principles for neural network design, development and testing, at determined levels of rigour. This would not be complete or uncontroversial, since there are radically different opinions about the relative merits of the fundamental statistical principles underlying not just neural networks, but also linear models, as well as advanced non-linear statistical models that are also making their way into practical applications. Nevertheless, at least the boundaries of applicability of different methods can be agreed for the most commonly used architectures.

It is worth noting that the international safety standard IEC 61508 [11] that aims to promulgate an international consensus on best practice currently discourages the use of neural network technology (and more generally, knowledge-based computing and artificial intelligence) in safety-related applications. Progress on the applicability of different methods for safety would be a valuable addition to the IEC 61508 approach. A holistic view of the neural network element in the overall safety-related system would put the neural network issues into better context.

In dynamic control and monitoring, there is less consensus of what constitutes a good model, but the current practice is much better informed about the design of purpose-built systems.

Best practice is not established to the same level of detail for software that adapts over time, such as Adaptive Resonance Theory and reinforcement learning, both already successful in routine practice for large-scale applications. However, the basic principles for best practice in assured design for safety-related applications would apply equally well to them.

Embedded systems emulating neuro-physiology, and other hardwired neural networks, are likely to fall more clearly within the scope of current hazards analysis frameworks.
Current practice and the need for best practice are addressed directly in a limited questionnaire, which I carried out to sample the community of neural network practitioners. The respondents are from Europe and the US, and include members of well-known user groups including the Neural Computing Applications Forum (http://www.ncaf.co.uk/) and NEuroNet (http://www.kcl.ac.uk/neuronet/). A summary of the replies received is listed in the Appendix.

The wide range of applications reported in the questionnaire replies is complemented with a an equally range of basic methodologies, from which the two examples of good practice quoted earlier, Bayesian approximation and Monte Carlo integration, are notable omissions. This indicates that it may be the whole engineering design, rather than the neural network modules, that are key in achieving engineering success.

A range of hazards is reported, including standards systems failures that are critical especially for operation in real-time, including hardware failure and system integration. Limited evaluation, failure to protect against novel data and a black-box approach clearly could be addressed in best practice guidelines.

There was some consensus that time-to-market can override formal systems design procedures, but this is not likely to be any more preponderant in neural that in general computing. Good practice will always have to defer to specific requirements for particular applications, therefore should not be prescriptive, but serve an educational and advisory rôle. Its 'teeth' could be in the award of 'quality seals', similar to the EuroSeal proposed by the TEAC study [8].

Prescriptive regulatory practices at the unit level for neural networks would be seen as premature and counter-productive to the technology transfer effort already put into small companies. The largest chorus of response was that sound practice in safety-related applications needs further guidance, and that it would encourage confidence in the technology, lack of which has been a barrier to entry for many companies.

Arguably the most important safety outcome overall would be raised awareness of safety, analogous to that achieved for PLCs with the Safety-Critical Systems Club.

6.5 IS THERE A NEED FOR RESEARCH TO IMPROVE BEST PRACTICE?

There is a need for further research to provide specific guidance concerning acceptable practice across the range of methodologies used for safety-related applications of neural computing. This can be addressed at each of three levels of design.

At the unit level, best practice guidelines are needed to update and extend those produced by the Neural Computing Technology Transfer Programme [9], since this is a rapidly developing field where the transfer of technology to market can be remarkably quick. The evidence from the applications reviewed in this report is that safety-related systems with neural network components use a small number of pattern recognition algorithms, and a range of purpose-build non-linear models for identification and control. Current best practice for them needs to be systematised.

Additional research is needed to bring a safety perspective into some of the new trends in neural network research that are becoming more widely used, often propagated via freeware. Beyond supervised models, they include non-linear visualisation tools based on self-organising maps, support vector machines and others based on formal computational learning theory, and unified frameworks for generative latent variable models, including Bayesian graphical models. This involves also the use software components downloaded from the web, with the consequent uncertainty in the functionality and performance of the component.

At the systems level, there must be recognition that inference systems form a continuum and do not operate in isolation. A consistent safety view must be therefore taken across the full range of system types, and also take account of radically new methods of delivery of computational intelligence, including remote electronic access.

Further research can build upon initial research on the regulation of intelligent systems, from the UK and elsewhere. This includes 'Guidance for FDA reviewers and industry' [4], the European project 'Towards evaluation and certification of telematics services for health' (TEAC-Health) [8], and the formal hazard schemes reviewed in the final report on 'Advances in safety critical systems: results and achievements from the DTI/EPSRC R&D programme' [10].

7 KEY CENTRES OF EXCELLENCE

Table 2.
Centres of excellence in the UK

Centre	URL	Themes
Neural Computing Research Group, Aston University	http://www.ncrg.aston.ac.uk/	Empirical modelling
MRC Biostatistics Unit, Cambridge University	http://www.mrc-bsu.cam.ac.uk/	Bayesian graphical modelling for medical applications
Artificial Intelligence Applications Institute, Edinburgh University	http://www.informatics.ed.ac.uk	Standards for knowledge management
Computer Science, Exeter University	http://www.dcs.ex.ac.uk/	Reliability of neural network systems
Computing and Mathematical Sciences, John Moores University	http://www.cms.livjm.ac.uk/research/snc/neural.htm	Industrial applications of neural networks
Centre for Systems and Control, Glasgow University	http://www.mech.gla.ac.uk/	Intelligent control systems engineering
NeuroNet, Kings College London	http://www.kcl.ac.uk/neuronet/	European network of excellence on neural networks
Adaptive Systems and Interaction, Microsoft Research	http://www.research.microsoft.com/adapt/	Modelling uncertainty
Neural Computing Applications Forum	http://www.ncaf.co.uk/	Industrial neural network applications
Chemical and Process Engineering, Newcastle University	http://www.ncl.ac.uk/chemeng/	Intelligent process control and monitoring
Department of Engineering Science, Oxford University	http://www.eng.ox.ac.uk/	Intelligent systems engineering
Computer Science, Sheffield University	http://www.dcs.shef.ac.uk/	Reliability of neural network systems
Electronics and Computer Science, Southampton University	http://www.isis.ecs.soton.ac.uk/	Intelligent systems and control
Knowledge Management Centre, University College, London	http://www.ucl.ac.uk/spp/hkmc.htm	Certification requirements for intelligent systems

Table 3.
International centres of excellence

Centre	URL	Themes
Center for Devices and Radiological Health, Food and Drug Administration, Washington DC, USA	http://www.fda.gov/cdrh/	Regulation of intelligent systems in medical devices
Software Research Laboratory, NASA, West Virginia, USA	http://research.ivv.nasa.gov/	Verification and validation of soft computing
Pacific Northwest National Laboratory, Washington State, USA	http://www.pnl.gov/	Industrial neural network applications

8 CONCLUSIONS

Neural network products are actively marketed, and some of them are routinely used in safety-related areas, including cancer screening and fire detection in office blocks. Safety benefits are claimed from improved performance, for example better specificity in alarms, or from a better theoretical framework in the design of controllers for non-linear environments, as in drive-by-wire. In the process industries there is potential for closer plant surveillance and consequently productive maintenance, including plant life extension.

Good practice was observed in the successful commercial applications reviewed in this report. Therefore, it is feasible to publish guidelines that will educate non-specialist users, and inform what is already a wide base of practitioners. However, there is no generally accepted theory to specify the body of evidence required to support performance claims for non-linear inference models, whether founded on symbolic reasoning, or non-linear statistical methods of any sort. By comparison, linear statistical methods have theoretical distributions for the inference error, against which claimed performance may be tested, for instance through randomised critical trials. This raises a significant technical barrier to the introduction of innovative intelligent systems into applications where, although financial and safety gains are potentially substantial, the limitations of existing theory mean that a regulatory obligation to demonstrate safety is difficult to meet in a cost-effective manner.

It is also clear from the applications reviewed that the key to successful transfer of neural networks to the marketplace is successful integration with routine practice, rather than optimisation for the idealised environments where much of the current development effort takes place. This requires the ability to evaluate their empirically derived response using structured domain knowledge, in addition to performance tests. In controller design, the scalability of solutions to production models, and the need to maintain safe and efficient operation over time, have led to the integration of linear design methods with neural network architectures.

Further research is necessary in two directions. One is to systematise current best practice in the design of a wide range of quite different neural computing software models and hardware systems. The other is to formulate a unified framework of best practice in high-complexity computing in safety-related applications, which is able to give a balanced perspective of the multiplicity of methods used, across the continuum from signal processing to predicate logic. Such a framework would help developers to construct persuasive safety arguments for advanced computing in safety-related applications.

There is an opportunity for voluntary guidelines on best practice, backed-up by an assessment of product design by a panel of experts acting for a suitably respected body (e.g. a trade association, or a DTI research club). A quality seal can be awarded as recognition of good practice, and to indicate a commitment, by product developers, to report on the practical utilisation of the product, including adverse events. This could be combined with a safety-related engineering AI club, to raise safety awareness among practitioners.

The availability of cheap and powerful computing, has promoted a fast expansion in artificial intelligence applications, where neural networks have a significant part. Some of the applications reviewed are safety-related, while others are indicative of future developments that are likely to involve critical aspects of control and monitoring, in a wide range of industries. There is still much work to be done before guidelines for intelligent systems meet the stern requirements of the "Scott-Ram test" (http://www.hse.gov.uk/dst/power/powrco11.htm), but it is clear that in the future new forms of computing will have as much of an impact on safety, as the introduction of electrical circuits.

DEFINITIONS AND ACRONYMS

Table 9-1 Definitions	
Neuromorphic	Computer systems that emulate neuro-physiological circuits
Regularise	Stabilise data fitting by explicitly penalising model complexity, usually with an additional term in the objective function
Sensitivity	True positive detection rate
Specificity	True negative detection rate
ROC	Receiver Operating Characteristic, a formalism to analyse detection rates across all values of the detection threshold
Self-organised model	Data density model commonly used for clustering and visualisation
Supervised model	Empirical model that attempts to replicate decision labels

Table 9-2 Acronyms	
AI	Artificial Intelligence
ART	Adaptive Resonance Theory: Neural network capable of incremental learning of class prototypes
BCS	British Computer Society
FDA	Food and Drug Administration
FDI	Fault Detection and Identification
FFT	Fast Fourier Transform
GTM	Generative Topographic Mapping: self-organised neural network with a noise model
GUI	Graphical User Interface
MLP	Multi-layer Perceptron: the most commonly used neural network for static pattern recognition
PID	Proportional, integral and derivative term: refers to process control
SOM	Self-organised mapping: neural network used to map high-dimensional data into usually two dimensions, by covering the data with a 2-D flexible surface then projecting the data onto it.
SWARM	Freeware for optimisation by simulating a large number of interacting software entities, called agents. Named by analogy to insects swarming a target and gradually finding an optimal path to it.
SVM	Support Vector Machines: decision model based founded on computational learning theory.

10 REFERENCES

1. Lisboa, PJG (ed.) ‘Current applications of neural networks’, Chapman and Hall, 1992.

2. Lisboa, P.J.G., Ifeachor, E.C. and Szczepaniak, P.S. (eds.) ‘Artificial neural networks in biomedicine’ Springer, 2000.

3. Lisboa, P.J.G., Vellido, A. and Edisbury, B. (eds.) ‘Neural networks applications in business’ World Scientific, 2000.

4. Guidance for FDA reviewers and industry, Office of Device Evaluation, US Department of Health and Human Services. http://www.fda.gov/cdrh/ode/57.html.

5. Proc. Committee on Computing, Information and Communications Workshop, Aug. 6-7, 1997, National Science and Technology Council of the USA http://www.ccic.gov/pubs/hcs-Aug97/hcs_problem_views.html.

6. Davies, M. and Owen, K. ‘Complex uncertain decisions: medical diagnosis’, Case Study 10 in Expert System Opportunities from the DTI’s Research Technology Initiative, HMSO 1990.

7. Szczepaniak, P.S., Lisboa, P.J.G., and Kacprzyk, J. (eds.) ‘Fuzzy systems in biomedicine’ Springer-Verlag, Berlin, 2000.

8. Forsström, J., *et al* ‘Towards evaluation and certification of telematics services for health (TEAC-Health) - key recommendations’, Health Telematics Application Programme HC 4101: Deliverable 8.1, University of Turku, Finland, 1999.

9. DTI Final Report ‘Evaluation of the parallel processing and neural computing applications programmes’ Assessment Paper 38 URN 99/922, 1999. (http://www.dti.gov.uk /tese/aurep38b.html).

10. Falla, M. ‘Advances in safety critical systems: results and achievements from the DTI/EPSRC R&D programme’, June 1997, http://tina.lancs.ac.uk/computing /resources/scs.

11. International Electrotechnical Commission ‘IEC 61508: Functional Safety of electrical/electronic/programmable electronic safety-related systems’. 1998.

11 APPENDIX. SUMMARY OF QUESTIONNAIRE REPLIES

The co-operation of researchers from the following organisations is gratefully acknowledged:

Aston University
The Boeing Company, Seattle, Washington
DuPont Central R & D, Wilmington, Delaware
Exeter University
Huddersfield University
Lund University
National Power
NCAF
NEuroNet, UCL
Nokia
Paisley University
Plymouth University
Politecnico di Milano
Southampton University
Sunderland University
Technical University of Lodz
UMIST

Q1. SAFETY-RELATED APPLICATIONS OF NEURAL NETWORKS

Q1.1. Are you aware of safety-related systems with neural network components ? (Y/N)

Yes:10
No: 8

Q1.1.1. Power generation and transmission
2/16 safety-related applications
3/15 not safety-related applications

Q1.1.2. Manufacturing industry
1/16 safety-related applications
4/15 not safety-related

Q1.1.3. Heavy industry (petro-chemical , marine, off-shore)
None reported

Q1.1.3. Transport (railways, automobiles, aircraft)
5/16 safety-related applications
1/15 not safety-related

Q1.1.4. Healthcare
7/16 safety-related
1/15 not safety-related

Q1.1.5 Consumer products
2/15 not safety-related

Q1.1.5.Other
1/16 safety-related
4/15 not safety-related

Q1.2 What is the role of the neural network component in the complete system?

Q1.2.1. Sensing
4 applications

Q1.2.2. Control
5 applications

Q1.2.3. Fault detection & identification
4 applications

Q1.2.4. Alarms
2 applications

Q1.2.5. Monitoring & off-line modelling
7 applications

Q1.2.6. Instrument calibration & self-checking
2 applications

Q1.2.7. Diagnosis
***8* applications**

Q1.2.8. Planning and scheduling
2 applications

Q1.2.9. Supporting high level human decision support
10 applications

Q1.2.10. Knowledge discovery and scenario analysis
3 applications

Q1.3. Extent of practical use

Q1.3.1. Routine use
10/31 applications

Q1.3.2. Occasional use
3/31 applications

Q1.3.3. Prototype tested but not in regular use
11/31 applications

Q1.3.4. Used only in process simulations
7/31 applications

Q1.3.5. Maintenance and up-grading protocol available
None reported.

Q2. THE STATE-OF-THE-ART IN INTEGRATED SYSTEMS WITH NEURAL NETWORKS

Q2.1. Methodology used to design the neural network

Q2.1.1. Vanilla back-propagation, or another standard predictive algorithm
9/31 applications Comment: this normally involves regularisation schemes

Q2.1.2. Regularisation schemes, Bayesian framework or Markov Chain Monte Carlo integration
None reported.

Q2.1.5. Error prediction in regression, marginalisation of the posterior in discrimination
1/31 application

Q2.1.6. Self-organised map
6/31 applications

Q2.1.7. Different networks in combination, and ensembles.
4/31 applications

Q2.1.8. Purpose-built machine learning algorithm
1/31 application

Q2.1.9. Interface with other approaches (e.g. fuzzy, expert systems, traditional controllers)
9/31 applications

Q2.1.10. ART1
1/31 application

Q2.2. How would you rate the dependence of safe operation of the system on the neural network component ?

Q2.2.1. Essential for correct operation of the system.
5 responses

Q2.2.2. Important but the system will operate successfully even if the neural network fails
3 responses

Q2.2.3. The risk associated with the neural network is minimal
2 responses

8 respondents did not report any safety-related applications of neural networks.

Q2.3. Where do you see the greatest potential hazard in the use of neural networks in practical systems ?

Q2.3.1. Errors arising from the complexity of the software implementation of the network
Yes: 2

Q2.3.2. Implementation errors arising from hardware failure (including failure to meet real-time constraints)
Yes: 2
No: 1

Q2.3.3. Mismatch between the network performance and pre-set specifications
Yes: 3

Q2.3.4. Limited performance evaluation on-line or in field trials
Yes: 7

Q2.3.5. Failure to protect against novel data
Yes: 12

Q2.3.6. Failure to verify the network operation against prior knowledge, e.g. by using explanation facilities
Yes: 4

Q2.3.7. Difficulties in integrating neural networks with other system components
Yes: 3

Q2.3.8. Comments
A big danger is ad-hoc approaches to NN training and acceptance of the view that a trained NN is a black box. Another danger may be that some people view them as a panacea.

Q3. NEED FOR DEVELOPMENT OF BEST PRACTICE

Q3.1. Are you satisfied with the extent of good practice as currently implemented in the practical neural network applications of which you are aware ?

Yes: 2
No : 2

Q3.1.1. Good practice in neural networks is comparable with alternative technologies (e.g. linear statistics, conventional process control, etc.)
Yes: 5
No: 2 Comment: good practice in NNs means underpinning technology with formal (statistical, probabilistic) models of the training data coupled with similar formal models of how particular technology applied (e.g. MLPs) interacts with data model.

Q3.1.2. Many applications are more concerned with time-to-market than with good practice.
Yes: 5 Comment: I'm sure that this is true on a scale way beyond NNs
No: 1

Q3.1.3. What is regarded as good practice in most applications I know, I would consider to be safe/unsafe
Comment: What is good practice is yet to be determined. It's currently a research question. However, industrialists claim that their product design follows good practice before they are used for real.

Q3.1.4. Comments
There is still something of a naïve approach in some applications, although these are usually less successful than those that know what they are doing. The breadth of material in the NN community may be too daunting for applications-oriented people to find what they need, so they may not know whether there is anything regarded as "good practice". The appears to be a need for guidelines, rather more technically-minded than any that are currently available.

Q3.2. Is there a need to promulgate official standards of good practice ?

Yes: 2

Q3.2.1. Good-enough practice still needs to be developed for neural networks
Yes: 3

Q3.2.2. Good practice exists and there is little risk in leaving neural network users long-enough to find it out
Yes: 2 Comment: Not just risk, but wasted time and opportunity),
No: 2

Q3.2.3. Good practice needs to be agreed and guidance needs to be published more widely
Yes: 11

Q3.2.4. Good practice in practical applications needs to be enforced somehow
Yes: 7 Comment: several respondents replied that guidance would be better than enforcement.

Q3.2.5. Comments
More research is needed. Meanwhile, engineers need to be involved in those special cases when a network is under consideration for deployment - expertise is need to verify and validate neural networks for each specific application.

Q3.3. How much is there to be gained, or lost, by enforcing good practice ?

Comment: 'good practice' is not perfectly definable, and will exclude some very good work and include some bad work.

Q3.3.1. Enforcement of good practice now is premature
Yes: 3

Q3.3.2. Enforcement now could deter further development of practical applications
Yes: 2
No: 1

Q3.3.3. Enforcement of sound practice in neural network design should already be in place where they are used for safety-related applications, and this is not yet done
Yes: 8

Q3.3.4. Poor use of neural networks in practice is lending the whole field a bad name
Yes: 11 It seems that a lack of clear good practice is a barrier to entry to the field in some cases.

Q3.3.5. Comments
Good practice should be enforced as in any other safety critical area.

Printed and published by the Health and Safety Executive
C1 03/01

HSE BOOKS

MAIL ORDER
HSE priced and free
publications are
available from:
HSE Books
PO Box 1999
Sudbury
Suffolk CO10 2WA
Tel: 01787 881165
Fax: 01787 313995
Website: www.hsebooks.co.uk

RETAIL
HSE priced publications
are available from
good booksellers

HEALTH AND SAFETY ENQUIRIES
HSE InfoLine
Tel: 08701 545500
or write to:
HSE Information Centre
Broad Lane
Sheffield S3 7HQ
Website: www.hse.gov.uk

CRR 327

£10.00

ISBN 0-7176-1971-0

9 780717 619719