CrossMark

Bio-inspired computing tools and applications: position paper

Tom Gedeon[1]

Abstract The confluence of significant computational power and inexpensive sensors provides the opportunity to reliably collect large volumes of information from the world and extract humanly useful information resources. This paper reviews a coherent body of work over the last 20+ years focused on development of advanced bio-inspired computing techniques, and their applications primarily for human related data in behaviour and human centered computing. We close with a synthesis proposing an experiment analysis methodology combining these tools.

1 Introduction

In the past twenty years, many bio-inspired computing techniques have been developed on toy datasets and applied to expensive small real datasets. Yet each year, our computers get faster, and we have access to more and more data collected from the real world. In the past decade we have seen that our access to computational resources has increased faster than the data we would like to analyze. An example of this is the success of deep learning models in image processing. The convolutional neural network (CNN) architecture was introduced in 1989 [1] yet it was from approximately 2006 onwards that their use in image processing became widespread.

This paper reviews a body of work from the past 20 or so years, focused on the initial development of techniques and some examples of their later application to human derived signals. The literature of such work is huge, this paper concentrates largely on the outputs of the late Information Engineering group at the University of New South Wales in Sydney from the 1990s (mostly Bio-inspired computing technique development) and later work up to the present in the Human Centered Computing research group at the Australian National University with applications and some further technique development. The objective is to provide a longitudinal survey which attempts to retain some breadth in topic areas as well as depth in some topics, and does so by the artifice of a group-biographic survey. The paper closes with a proposal for automated data analysis which synthesizes the bio-inspired tools discussed into an automated experiment analysis methodology.

2 Bio-inspired Computing Tools

Bio-inspired computing is related to artificial intelligence, machine learning and so on. The key is the inspiration for learning algorithms from the world, particularly from biology. In this paper we will concentrate on three main models, neural networks (including deep learning), fuzzy logic, and evolutionary algorithms. We note that the terms "soft computing" and "computational intelligence" are near synonyms of bio-inspired computing. Neural networks are a computational model based on models of neurons in

✉ Tom Gedeon
tom.gedeon@anu.edu.au

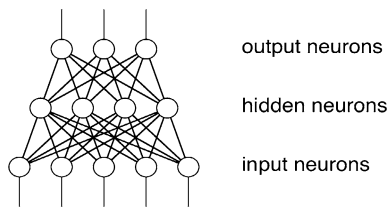1   Research School of Computer Science, The Australian National University, Canberra, Australia

🖄 Springer

8

Int. j. inf. tecnol. (March 2017) 9(1):7–17

Fig. 1 Simple neural network, multiple layers of hidden neurons are possible. If there are 'many' this becomes a deep learning neural network

the brain, and the way these simple computing elements are interconnected to produce complex computations. In the literature they are sometime referred to as artificial neural networks. It is particularly worth noting in the context of neural networks that these are computational (or engineering style) approximations and simplifications which have been found to be useful computationally, but do not attempt to directly model the real behaviour of neurons and biological nervous systems. Neural networks are in general local search techniques. Fuzzy logic can be seen to model linguistic reasoning, with multi-valued set memberships and linguistic variables. Evolutionary algorithms mimic some of the properties of biological evolution to perform computations in a form of randomized but guided global search. The combination of local and global search is slow, but can provide results better than either, in hybrid algorithms.

In the following sections we briefly review some work in these areas, some of which we believe can and should have application to modern large problems. Our confidence comes from two sources, firstly the success of the CNN architecture, and the recent implementation of the "Gedeon method" [2] in the H2O [3] deep learning package – currently still somewhat slow when applied to large datasets yet its implementation indicates an independent confidence in both the usefulness of the technique and its likely applicability on tomorrow's faster hardware.

2.1 Neural networks

The back propagation neural network algorithm was introduced in 1986 [4] (see Fig. 1).

The neural network is trained by input of training patterns, then the weights on each link are used to modulate the inputs, these are summed at the hidden neurons and a non-linear activation function is applied (often the sigmoid or logistic function). When values reach the output neurons, the value is compared to the desired value for this training input, and the difference serves as a 'back propagated' error signal to make small modifications to the weights in the preceding layer. These modifications can be used to infer error signals to earlier layers of weights

working back towards the input. After much presentation of training data, the network approximates the function linking the input to the output. This is an expensive process, particularly if the input is large and the network has many layers. As mentioned earlier, it is only in the last decade that deep learning on large data has become accessible to those without a handy supercomputer.

2.1.1 Gedeon method

Data encoding and feature selection for the training of back-propagation neural networks has two basic principles: i) to avoid encrypting the underlying structure of the data, and ii) to avoid using irrelevant inputs. The paper [2] used weight matrix analysis and functional measures on two noisy real data sets, and a novel aggregation technique was introduced

$$P_{jk} = \frac{|\mathrm{w}_{jk}|}{\sum_{r=1}^{nh} |\mathrm{w}_{rk}|}. \tag{1}$$

where P is the contribution of a hidden neuron to the output. This can readily be calculated for contributions of the input to the first hidden layer, or composed backwards for any number of layers, for example for a 2 layer network:

$$Q_{ik} = \sum_{r=1}^{nh} (P_{ir} \times P_{rk}) \tag{2}$$

2.1.2 Bimodal distribution removal

Cleaning noisy training sets can improve generalization, but many methods perform well on artificially noisy data, but not so well on real world data where distinguishing between rare data points and just noisy data can be difficult. A statistically based method [5] performs well on such data and provides a stopping criterion to terminate training (see Fig. 2).

2.1.3 Adding noise

While we have access to more and more data, sometimes we are still short of labeled data, where the class label or quality value and so on is assigned using an expensive process, such as expert human intervention or the collection process of the raw data is itself expensive, such as extraction petroleum reservoir core samples or values for a geographical information system 'pixel'. Experiments have shown that a crude form of simulated annealing with decreasing amounts of noise works well [6], and that varying amounts of normally distributed noise also leads to good results [7].

Fig. 2 Error distribution by input patterns, a few epochs training (*left*) and 500 epochs (*right*)
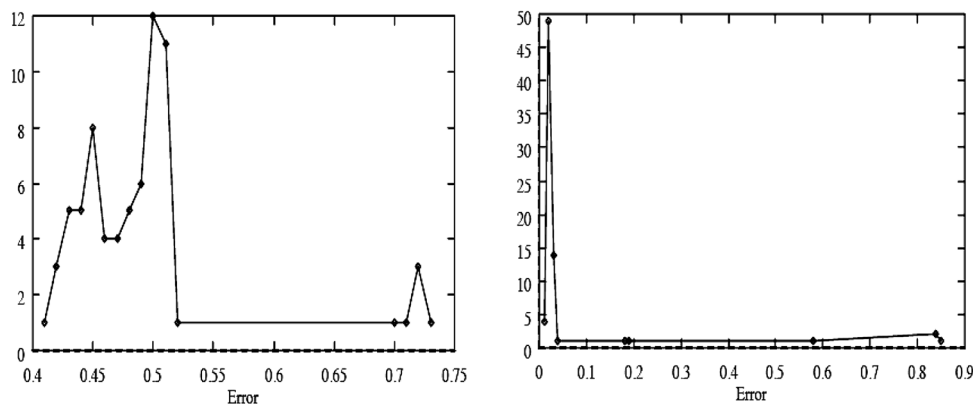
2.1.4 Explanations

The use of neural networks for prediction is commonplace, for example in student grade prediction [8], but in domains of human interaction, this is unsatisfactory. Thus, for example, if we predict the weather tomorrow, we expect this to have limited effect on the weather. Yet if we predict a student's grade, we expect some action to (attempt to) change that prediction if the student is unhappy with the predicted result or even the converse may occur when a feeling of complacency may lead to less effort and a lower mark. Producing explanations based on causal modeling [9] can produce explanations for the neural network conclusions, noting that correctness here is in matching the network outputs and not the underlying class distribution, a distinction which will become significant later in this paper. We note that the approach was expensive computationally, and was applied to clustering based subsets of the data, and produced rules of the form (see Fig. 3):

We note that the negative association of *h2* with a Distinction grade proved correct. There is recent related work in peer marking [10].

2.1.5 Cascade networks

The cascade correlation network [11] is a powerful training algorithm which constructs networks one neuron-layer at a time. Each new neuron has connections from the inputs and all previous neurons which allow arbitrarily complex learning in the last neuron in principle. In practice, to keep computational costs manageable, each neuron-layer's weights are frozen before the next is added. The networks therefore freeze inaccurate early learning and require many layers to unlearn such early learning, causing these networks not generalize well on regression and some classification problems. An alternative approach using RPROP to train the networks and low learning rates (to reduce weight updates rather than freeze weights), results in networks which use fewer hidden neurons and generalize better than those produced by the original cascade correlation algorithm [12]. An extension to insert small 'towers' of cascaded neurons as single higher order neurons reduced the computational cost close to (then) tractability – we are investigating these as deep learning feature extractors for non-visual data (see Fig. 4).

2.1.6 Extreme Learning Machines

The extreme learning machine (ELM) is a method of training shallow feed forward networks with high efficiency [13]. The key notion is that input weights are not trained, rather they are set at random and fixed. Then, since the training set and input weight matrix is fixed, the output weights can be calculated directly, generally by the Moore-Penrose pseudo-inverse. The number of hidden neurons need to be raised significantly, yet an increase from 20

Characteristic pattern	Rule Set
ON Distinction	(h2 ≤0.52) and (mid ≤0.98)
ON Credit	(Crse ≥092) and (Stage ≤0.93) and (mid ≥0.12) and (lab10 ≥0.14)
ON Pass	mid ≥0.85

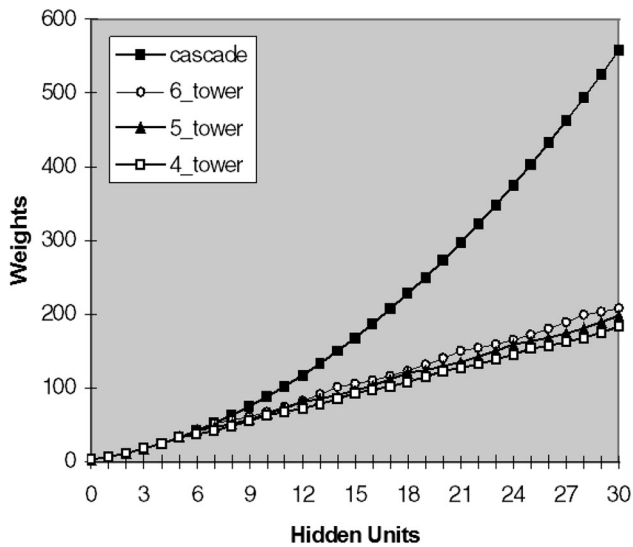**Fig. 3** Explanations for a Distinction grade using causal index with characteristic patterns

Fig. 4 Exponential growth in weights is reduced using towers – higher order neurons

Fig. 5 Cascade of ELM networks

hidden neurons in back propagation can be replaced by 400 neurons with ELM training and still achieve a 10-15 fold increase in speed, with small drop in accuracy. Each ELM hidden neuron can be considered to come with some initial (random) fixed functionality and the training of the output weights is a weighted selection from the available menu. A small ELM network could be used as a higher order neuron, and some initial work has been done [14] (see Fig. 5).

Such sequential processing may be useful in simpler settings, where two outputs are to be predicted, but the prediction of one when added as an extra input can improve the prediction of the second [15].

2.2 Fuzzy logic – Fuzzy relational maps, interpolation, hierarchical fuzzy, FOE

The term 'fuzzy' is unfortunate; the name intended to signify a mathematically rigorous method to deal with uncertainty. Unfortunately, the normal English meaning of the word is close to the opposite of this. Thus, much of the initial successes in this field originate in non-English speaking countries, particularly Japan.

Fuzzy logic [16] approximates human linguistic reasoning. When asked to define 'tall' we can clearly identify both tall and short people, yet naturally do not consider the two individuals on either side of an arbitrary height to be unambiguously tall or short when their heights may differ by only a few millimeters. Thus, values of set membership between 0 and 1 are used. These latter individuals may have fuzzy set membership values of 0.49 and 0.51 reflecting their similarity in height and retaining the ability to convert to classical sets by the round function (also called defuzzification). The second main component is
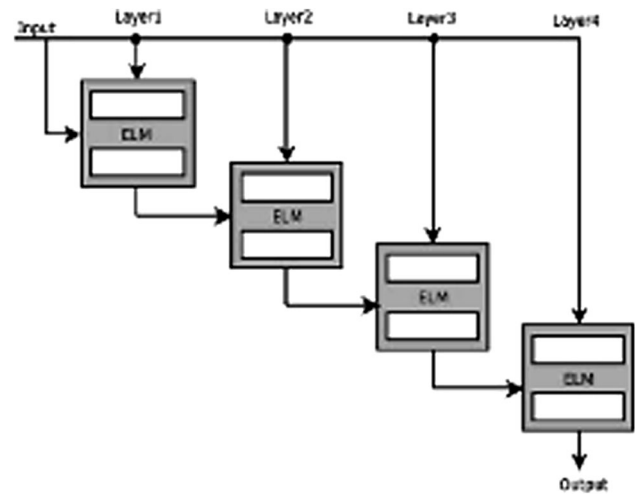
fuzzy linguistic variables, such as the example of 'tall' used already.

In the previous section, neural networks return values between 0 and 1, as do fuzzy sets, and probability. It is important to differentiate between these. A neural network's outputs are not probabilities unless the network has been explicitly trained with output values representing probabilities [17]. The values between 0 and 1 in fuzzy logic represent vagueness and are possibilistic, unlike probabilistic values of 0 to 1 which model ignorance. A simple dichotomy: probabilities add to 1, while possibilities need not do so, though for convenience many fuzzy algorithms impose the same condition. Figure 6 demonstrates the former, we not the example shown has values of 0.4 and 0.15 which add to less than one, while the converse is also possible. A key limitation of fuzzy logic is the exponential growth in number of rules for full (dense) rule bases. This had limited applications to control settings where there can be only a few input parameters but with complex behaviour. See equation 3, where k is the number of input dimensions and T is the granularity of the rule base – the number of fuzzy linguistic terms per input dimension

$$|R| = O(T^k). \tag{3}$$

2.2.1 Fuzzy Equivalence

Fuzzy systems and neural networks are (most models) universal approximators in that they can approximating any continuous function to arbitrary accuracy. These techniques therefore share approximation capabilities. Fuzzy systems with if-then rules have an advantage of easy interpretability, and neural networks can adapt their learning to improve performance on a training data set. It has been shown that several fuzzy controllers implement
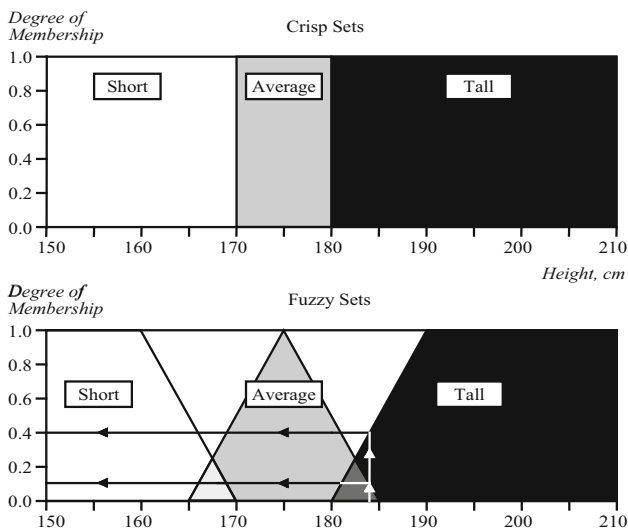
Fig. 6 Tradition ('crisp') sets and fuzzy sets

radial basis function neural networks [18], a kind of feed forward network with radial basis activation functions rather than the sigmoid ones described above. Such evidence suggests that combination of these techniques in hybrid systems will not lead to a loss of approximation ability.

2.2.2 Fuzzy Cognitive Maps

The FCM [19] is an extension of cognitive maps with directed links with weights which represent degrees of causality, which extends the notion of the traditional cognitive map decision-making aid. The causal links between concepts are straightforward for people to define, but produce static maps. The weighted links with a learning regime allows for modeling of dynamic situations. The learning process uses differential Hebbian learning and event sequences with an initial static map to learn the weights which best reproduce such training sequences [20].
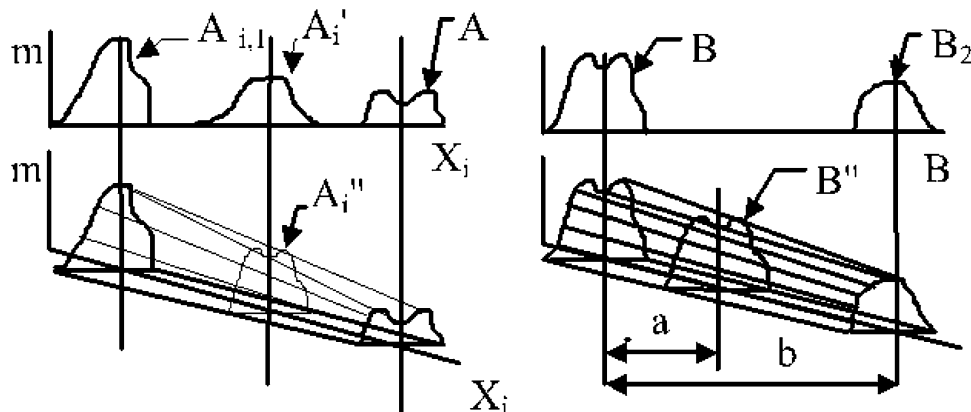
2.2.3 Qualitative Modeling

Sugeno and Yasukawa's (SY) qualitative modeling [21] creates a fuzzy rule base (fuzzy IF–THEN rules) from input–output data thus importing one of the benefits of neural networks, and assigns linguistic labels to fuzzy sets in the rule base. Subsequently, some properties such as the construction of trapezoid membership functions, rule projection from training data, selection of important variables and finally parameter identification were improved [22], as was the cluster search for rules projection [23].

One of the advantages of the SY method is that only the required rules are produced, and hence does not produce a dense rule base, potentially reducing the combinatorial explosion. A concomitant disadvantage is that if there are test samples with values 'between' the rules, the result is not guaranteed to be well defined, unless a sparse fuzzy technique such as fuzzy interpolation is used. Another approach to sparse fuzzy system generation is the use of clustering in the output space with projection to each input dimension to produce one dimensional input clusters which are merged to produce fuzzy rules [24].

2.2.4 Fuzzy Interpolation

These techniques are all descendants of Kóczy and Hirota's linear interpolation technique (KH interpolation, based on α-cuts) [25]. Fuzzy interpolation reduces the computational complexity by reducing T in equation 3. The basic notion of fuzzy interpolation is that antecedents (inputs) between two antecedent fuzzy sets of a pair of rules will produce consequents (outputs) between the consequent parts of these two fuzzy rules. This is plausible only if we assume there are no discontinuities and sudden changes such as in symbolic spaces. Since we use fuzzy logic to model the real world which is generally continuous in nature, this is a reasonable assumption (as it is for neural networks and for evolutionary algorithms).

Fig. 7 Determination of A_i'' and B'' by geometric solid cutting

KH interpolation had its drawbacks, which could lead to non-interpretable conclusions due to the way the two core points were interpolated for triangular observations with trapezoidal rules, in some cases. Further the technique required convex and normal fuzzy sets. There have been a number extensions, we survey some different approaches: the use of a spatial geometric representation allows us to avoid the requirements of convex or normal fuzzy sets, and guarantees interpretable conclusions in all cases [26] (see Fig. 7); an extension on the latter technique using the interpolation of the semantics of the rules and their inter-relationships to guarantee the direct interpretability of the conclusions and piecewise linearity for triangular membership functions [27]; finally, returning to modified α-cut interpolation (MACI) methods which retain the low computational complexity of the original KH method, firstly retaining vector description of the fuzzy sets as characteristic points, coordinate transformation, and considering the fuzziness flanking information in the input spaces at the conclusion leads to efficient interpolation of fuzzy rules for multidimensional input variables [28]; and a generalization of characteristic points for different a-cut levels with normalization and aggregation functions leads to always acceptable conclusions [29].

2.2.5 Hierarchical Fuzzy Systems

Another approach to reducing the exponential growth in number of rules is hierarchical structuring of the rule base so that only some rules are used at a time and the rules used reduce the choices of subsequent rules. That is, in effect reducing k in equation 3.

The first example of a hierarchical fuzzy system is the handcrafted rules for helicopter flight control [30]. A general problem in replicating that work has been that in many cases the border conditions between hierarchical branches require bridging rules which can often require many or all variables from both branches and hence obviating the benefit of the hierarchical system. The use of fuzzy relations [31] or co-occurrence relations [32] based on fuzzy tolerance (compatibility) relations and fuzzy similarity (equivalence) allows for extension of search based on hierarchical co-occurrence of words and short phrases in documents. This approach thus is based on the document structure and on the semantic interrelationships of words. Equivalent structuring information is not always or even often available. Of more general utility, hierarchical rule bases can be constructed directly from data [33]. This technique constructs simple single variable at each level hierarchical rule bases and then prunes the set of hierarchical fuzzy rules to directly reduce the rule base, see Fig. 8 – the 12 rules are reduced to 5.

When a whole sub-rule bases have the same output class (*) then it can be pruned by moving the class label up the hierarchy. In some cases (**) a value could be interpolated and also removed.
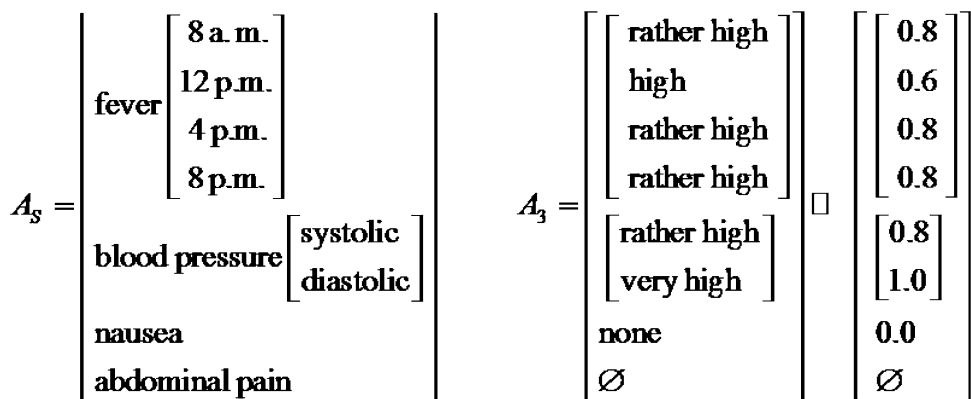
2.2.6 Fuzzy Signatures

Extending fuzzy sets to vector components provides some structuring advantages, but allowing components of the vector to be vectors themselves produces a tree structure which by its nature produces hierarchical structures [34] we can build ab initio. Figure 9 shows a possible data structure for a fuzzy signature based on doctors' assessments.

The example in Fig. 9 is illustrative, in a medical ward, instead of the qualitative components; the vector element might be *temperature* and the values to be specific measurements. In practice, such signatures are meant to be flexible, so how do we compare patient A_3 in Fig. 9 above, with patient A_4 (not shown) whose fever was just measured once? We do this by aggregating the values. Some possible simple functions could be the maximum value (so A_3's fever value is aggregated to 0.8), or the average, in some

Fig. 8 Pruning of simple hierarchical rule base from 12 to 5 rules

R1:
If Experience is Little then use R21
If Experience is Some then use R22
If Experience is Good then use R23

R21:
If Age is Young then Salary is Basic*
If Age is Middle then Salary is Basic*
If Age is Old then Salary is Basic*

R22:
If Age is Young then Salary is Basic
If Age is Middle then Salary is Fair**
If Age is Old then Salary is High

R23:
If Age is Young then Salary is High *
If Age is Middle then Salary is High *
If Age is Old then Salary is High *

R1:
If Experience is Little then Salary is Basic
If Experience is Some then use R22
If Experience is Good then Salary is High

R22:
If Age is Young then Salary is Basic
If Age is Old then Salary is High

Fig. 9 A basic fuzzy signature structure A_S and an instantiation A_3. Depending on diagnosis, the nausea element (say) could be a vector of measured values even hourly

$$A_S = \begin{bmatrix} \text{fever} \begin{bmatrix} 8\,\text{a.m.} \\ 12\,\text{p.m.} \\ 4\,\text{p.m.} \\ 8\,\text{p.m.} \end{bmatrix} \\ \text{blood pressure} \begin{bmatrix} \text{systolic} \\ \text{diastolic} \end{bmatrix} \\ \text{nausea} \\ \text{abdominal pain} \end{bmatrix} \qquad A_3 = \begin{bmatrix} \begin{bmatrix} \text{rather high} \\ \text{high} \\ \text{rather high} \\ \text{rather high} \end{bmatrix} \\ \begin{bmatrix} \text{rather high} \\ \text{very high} \end{bmatrix} \\ \text{none} \\ \varnothing \end{bmatrix} \square \begin{bmatrix} \begin{bmatrix} 0.8 \\ 0.6 \\ 0.8 \\ 0.8 \end{bmatrix} \\ \begin{bmatrix} 0.8 \\ 1.0 \end{bmatrix} \\ 0.0 \\ \varnothing \end{bmatrix}$$

cases the minimum and more complex functions are possible. In general we would like to learn the aggregation functions which best suit the data we have [35], which achieves good results in SARS diagnosis. Beyond the medical domain, fuzzy signatures have been used in robot communication with the introduction of a codebook to implement implicit fuzzy communication [36].

2.2.7 Fuzzy Output Error

The final fuzzy approach we will discuss is to use the notion of fuzzy values being possibilistic and about uncertainty in learning algorithms. That is, when output values are 'close enough' we should not be training to improve those output values but rather be training those which are not correct at all. This leads to the notion of fuzzy output error [37] which permits choices in the shape of error functions and allow us to tune error functions for specific applications. For example, in the medical domain false positives (has condition but algorithm classifies as healthy) are very unfortunate for serious conditions.

From Fig. 10 we can see that in general it appears sensible to include the fuzzy values of classification result values between 0.5 and 1 rather than just the rounded value. A sensible extension is to learn the shapes of the membership functions from data [38], using squashing functions with sigmoids approximating the fuzzy membership functions so gradient descent techniques such as neural networks can be used.

2.3 Evolutionary Algorithms and Hybrid Techniques

Evolutionary algorithms use fitness functions to determine the likelihood that an individual in the population of solutions is used to help produce the next cycle of candidate solutions. Operators such as cross-over and mutation mimic biological mechanisms which combine information from two 'parent' solutions to create a new candidate
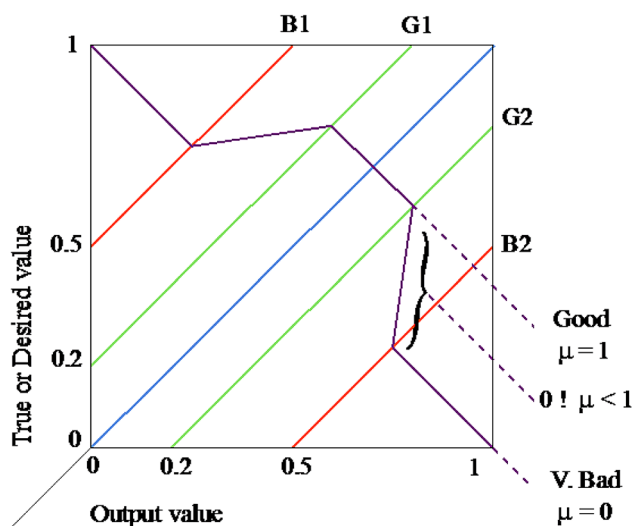

Fig. 10 Superimpose a fuzzy trapezoidal (or triangular) membership function on the cross diagonal for illustration purposes

solution and introduce diversity ('mutation' increases the possible search space, whereas 'cross-over' optimizes the search of the space already available to candidate solutions. At a practical level, if we have a set of labelled data we would initially consider a neural network solution, whereas if we have a function to evaluate the quality of solutions we would initially consider an evolutionary algorithm. Of course, functions can be used to generate data, and data sets can be used instead of a function by aggregating the errors and invert that cost function as a fitness function.

In principle, the individual elements of candidate solutions which together form the solution (and form the *chromosome*) should be independent and continuous, though in practice evolutionary algorithms are sufficiently robust to provide good solutions in without independence or continuity [40]. For example, timetabling is inherently a dependent problem as a change in one timeslot inevitably causes changes to some or all of the other timeslots [41]. In such situations clustering of the data [42] is useful to create independent groups of points.

A survey can be found in [43]. Hybrids of neural networks and evolutionary algorithms were mentioned earlier, other hybrids are possible, for example fuzzy and evolutionary [44], or ensembles [45].

3 Human Behaviour

The significance of human centered computing research rests on the ability to be useful in human contexts – to extract humanly useful information resources and use these to enhance computer and software interactions with human beings. We review a few approaches in this vein, ranging from automatic camera control based on human behaviour, face recognition, reading on large and small screens, and recognizing gendered differences in behaviors. This latter serves as a proxy for all of the other possible differences (education, culture, first language and so on) which lead to the need to consider sub-groups of people in such research.

3.1 Camera Control

The complexity of dealing with human beings in experiments is well illustrated by the study in which it was shown that the sequence of experiments has an effect on the results [46]. Normally, the purpose of multiple sequences is to determine whether there is a 'first seen' issue which benefits that technique. Subsequent work showed improvements in pan-tilt-zoom camera control from a natural combination of head movements and eye gaze [47], leading to a simple model where automatic eye movements of the operator are used to control the camera [48], or using mixed media [49]. These studies illustrate that application of advanced processing is best done behind the scenes and to aid users rather than force them to make choices or to learn new behaviors. With the computational power at our disposal and the bio-inspired computing tools previewed earlier, this could be done more broadly.

3.2 Face Recognition

For people, faces are important, thus a survey of this nature no matter how brief would be remiss without considering computational approaches. There are two aspects, firstly the recognition of faces varying in pose, illumination and expression [50]. The next steps are expression recognition, emotion recognition (including by other techniques [51], emotional state [52]), followed by group affect [53].

3.3 Reading and eLearning

Reading is a learnt behaviour and does not come naturally to humans. Nevertheless, due to the substantial amounts of time spent learning to read and practicing reading, it becomes close to a natural human behaviour for a large part of the world's population, except for some with special needs [54]. For business, each group of people can have specific needs to address [55].

During reading, we can use bio-inspired and information retrieval tools to discover information on the texts being read by observing and analyzing the readers' behaviour. We can extract the stress level embedded in the text scenario [56], the reading comprehension of the reader [57], and the document category [58].

On small screens, we can discriminate search behaviour [59] and mobile learning preferences [60]. For eLearning in general, we can discover how the mode of presentation of the texts read affects learning outcomes [61].

3.4 Gender

We can recognize differences in behaviour in reading by gender [62], and differentiate English as a first language readers from second or later language readers [61]. We can differentiate responses to face replacement in videos [63]. Such evidence supports our contention that results from human studies are contextual and we must use the large volumes of data available to ensure we cover the breadth of possible populations rather than collect ever more data from the same populations. As mentioned earlier, we consider the work on gender to be a general proxy for such considerations.

3.5 Stress

Stress is a major bane of modern society, so detection of stress warrants its own section. The literature is huge; we will mention a particular approach focused on the observers of stress. This makes those techniques suitable for the most common work and play setting of this era – looking at things on screens. We illustrate by an example using information not readily available to fellow humans, being high resolution thermal images [64, 65], see Fig. 11.

4 Synthesis – Deep Learning for Experiment Synchronization

Data structuring and synchronisation, a synthesis and proposal: we wish to record vast quantities of data from sensors attached to or pointing at human participants in experiment, and need to be able to reason and make comparisons between data recorded for different people using different sensors, in somewhat different settings. (Humans do this all the time!) We can use our *event signatures* (an extension of fuzzy signatures as described earlier) to record rich metadata along with the sensor recordings.
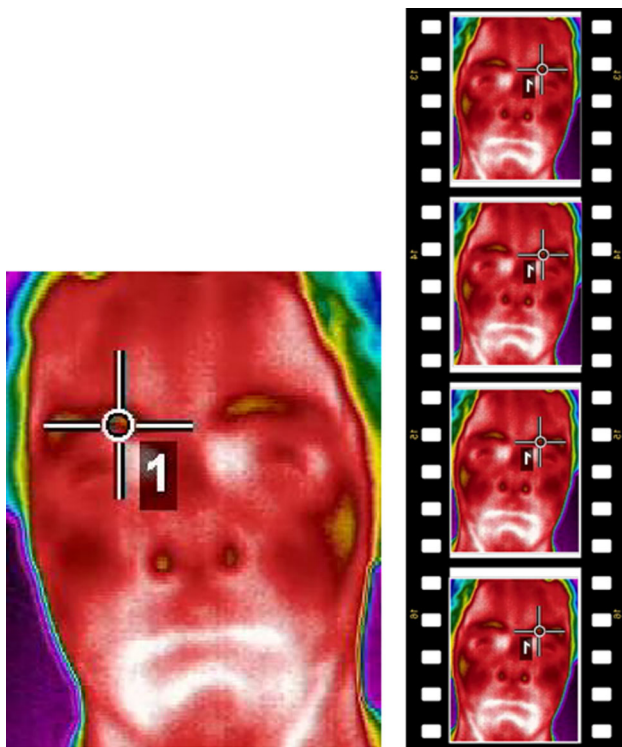
Fig. 11 Thermal video of experimental subject watching stressful / calm videos

convolutional neural networks). With our time sequence sensor data, we know which signals are from the same experiment and will always roughly know the time, so the task is to refine that (see Fig. 12) knowledge

We use the deep learning approach to discover the best possible match between sensor signals using the various sensors. We provide two simple examples here: 1) frontal EEG electrode signals always contain some noise to be filtered out, caused by signals from the eye muscles during saccades: this noise is well correlated with eye gaze data and is the between-fixation times of that signal [66]. The patterns of lengths of saccades can accurately synchronise these two signals. 2) The effect of breathing on heart rate is detectable from that signal, and can be matched to warmer pixels near the nostrils/mouth [67] thus synchronising those sensors. Risks, caveats and counters: a) we note that this adaptive synchronisation will not be millisecond precise, but neither are human reactions; b) it uses the overt surface statistical properties of the signals (such as long-range correlations in the signals; we note that the specific reactions within our experiments will not have the same long range correlations due to order balancing or experiment; we gave an example using the noise from EEG to correlate with the least interesting part of the eye gaze signal); and c) that this synchronisation is testable, as we know of the different timings of skin conductance reactions as compared to fNIRS or EEG reactions to the same event, so we would be able to calibrate our adaptive synchronisation.

The plethora of individual devices means strong synchronisation signals are not possible with all or even the majority of devices. We can solve this problem in two ways, simply storing sensor device times in the event signatures as differences in device time stamps are likely to persist, but our innovative solution is automated synchronisation and data alignment using deep learning with convolutional neural networks. Deep learning has shown great success in image analysis, devised originally for optical digit recognition where the 2D structure of the data is known (digits can be translated, and even rotated somewhat and still remain recognisable to humans and

5 Conclusion and Future Work

We have described a longitudinal body of work in the construction of bio-inspired tools, increasingly focused on applications to signals we can directly record or extract from human behaviour, and closed with a synthesis and proposal. A follow-up report on advances in human centered computing over the past few years will be our next project.

Fig. 12 Sketch of CNN: inputs are sensor signals, outputs are offset vectors
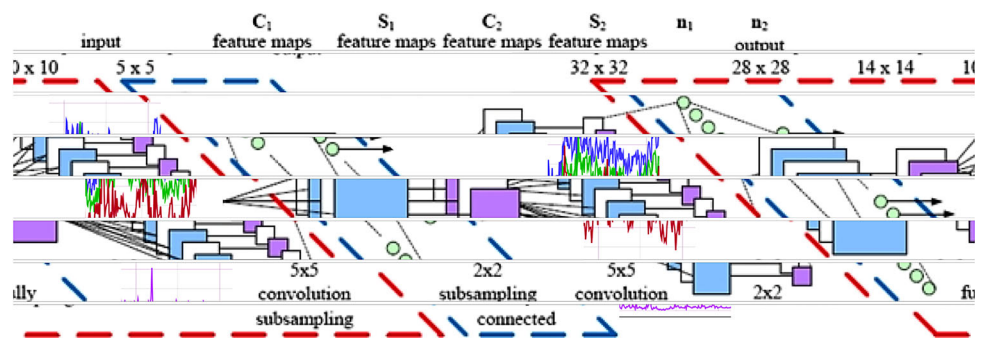
References

1. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1989) Back propagation Applied to Handwritten Zip Code Recognition. Neural Computation 1(4):541–551

2. Gedeon TD (1997) Data mining of inputs: Analysing magnitude and functional measures. International Journal of Neural Systems 8(2):209–218

3. Candel, A., V. Parmar, E. LeDell, and A. Arora. Deep Learning with H2O. H2O. ai Inc. www.h2o.ai (2016)

4. Rumelhart, DE, Hinton, GE, Williams, RJ, Learning internal representations by error propagation, in Rumelhart, DE, McClelland, Parallel distributed processing, vol. 1, MIT Press, (1986)

5. Slade P, Gedeon TD (1993) Bimodal Distribution Removal, in Mira, J., Cabestany, J. and Prieto, A., New Trends in Neural Computation, Springer Verlag. Lecture Notes in Computer Science 686:249–254

6. Bustos RA, Gedeon TD (1995) Learning synonyms and related concepts in document collections. Alspector, J., Goodman, R. and Brown, TX Applications of Neural Networks to. Telecommunications. 1995(2):202–209

7. Brown, W. M., Gedeon, T. D., & Groves, D. I. (2003). Use of noise to augment training data: a neural network method of mineral–potential mapping in regions of limited known deposit examples. Natural Resources Research, 12(2), 141-152. (2003)

8. Jishan ST, Rashu RI, Haque N, Rahman RM (2015) Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. Decision Analytics 2(1):1

9. Gedeon, T.D. and Turner, H. Explaining student grades predicted by a neural network, Proceedings International Joint Conference on Neural Networks, Nagoya, 609-612. (1993)

10. Caldwell, S.B. and Gedeon, T.D. Optimising Peer Marking with Explicit Training: from Superficial to Deep Learning, 1st International Conference on Higher Education Advances. (pp. 626-631). DOI:http://dx.doi.org/10.4995/HEAd15.2015.441 (2015)

11. Fahlman, S.E., and Lebiere, C. The cascade-correlation learning architecture. In Advances in Neural Information Processing II, Touretzky, Ed. San Mateo, CA: Morgan Kauffman, 1990, pp. 524-532. (1990)

12. Treadgold, N.K. and Gedeon, T.D. A Cascade Network Algorithm Employing Progressive RPROP, in Mira, J, Moreno-Díaz, R and Cabestany, J, (eds.), Biological and Artificial Computation: From Neuroscience to Technology, 733-742, Springer Verlag, Lecture Notes in Computer Science, vol. 1240. (1997)

13. Huang, G. B., Zhu, Q. Y., & Siew, C. K. Extreme learning machine: a new learning scheme of feedforward neural networks. In Neural Networks. Proceedings. IEEE International Joint Conference on (Vol. 2, pp. 985-990). (2004)

14. Gedeon T. and Oakden A. Extreme Learning Machines with Simple Cascades. In Proceedings of the 5th International Conference on Simulation and Modeling Methodologies, Technologies and Applications ISBN 978-989-758-120-5, pages 271-278. DOI:10.5220/0005539502710278 (2015)

15. Wong PM, Taggart IJ, Gedeon TD (1995) The Use of Neural Network Methods in Porosity and Permeability Predictions of a Petroleum Reservoir. AI Applications 9(2):27–37

16. Zadeh, L.A. Fuzzy sets. Information and Control. 8 (3): 338–353. Doi:10.1016/s0019-9958(65)90241-x. (1965)

17. White H (1989) Learning in artificial neural networks: A statistical perspective. Neural computation. 1(4):425–464

18. Koczy L, Tikk D, Gedeon T (2000) On functional equivalence of certain fuzzy controllers and rbf type approximation schemes. International Journal of Fuzzy Systems 2(3):164–175

19. Kosko B. Fuzzy cognitive maps. International Journal of Man-Machine Studies. 1:24(1):65-75. (1986)

20. Khan MS, Chong A, Gedeon TD (2001) A Methodology for Developing Adaptive Fuzzy Cognitive Maps for Decision Support. Journal of Advanced Computational Intelligence 4(6):403–407

21. Sugeno M, Yasukawa T (1993) A fuzzy logic based approach to qualitative modeling. IEEE Trans. Fuzzy Syst. 1:7–31

22. Tikk D, Bíró G, Gedeon TD, Kóczy LT, Yang JD (2002) Improvements and critique on Sugeno's and Yasukawa's qualitative modelling. IEEE Transactions on Fuzzy Systems 10(5):596–606

23. Wong, K.W., Kóczy, L.T., Gedeon, T.D., Chong, A., Tikk, D. Improvement of the Clusters Searching Algorithm in Sugeno and Yasukawa's Qualitative Modeling Approach. 7th Fuzzy Days in Dortmund - International Conference on Computational Intelligence, October 2001, Dortmund, pp. 536-549. (2001)

24. Chong, A., Gedeon, T.D. and Kóczy, L.T. Projection based method for sparse fuzzy system generation, Proceedings of WSEAS International Conference on Scientific Computation and Soft Computing, Crete, pp. 321-325. (2002)

25. Kóczy L, Hirota K. Approximate reasoning by linear rule interpolation and general approximation. International Journal of Approximate Reasoning. 1:9(3):197-225. (1993)

26. Baranyi, P., Gedeon, T.D. and Koczy, L.T. Rule interpolation by spatial geometric representation. In Proc. IPMU (Vol. 96, pp. 483-488). (1996)

27. Baranyi P, Mizik S, Kóczy LT, Gedeon TD, Nagy I. Fuzzy rule base interpolation based on semantic revision. In Systems, Man, and Cybernetics, IEEE International Conference (Vol. 2, pp. 1306-1311). IEEE. (1998)

28. Wong, K.W. and Gedeon, T.D. Petrophysical Properties Prediction Using Self-generating Fuzzy Rules Inference System with Modified Alpha-cut Based Fuzzy Interpolation, The Seventh International Conference of Neural Information Processing (ICONIP2000), November 2000, Taejon, 1088-1092. (2000)

29. Tikk D, Baranyi P, Gedeon TD, Muresan L (2001) Generalization of the rule interpolation method resulting always in acceptable conclusion. Tatra Mountains Mathematical Publications 21:73–91

30. Sugeno M, Murofushi T, Nishino J, Miwa H. Helicopter flight control based on fuzzy logic. In Proceedings of the International Fuzzy Engineering Symposium (IFES'91). pp. 1120-1121. (1991)

31. Tiwari RG, Husain M, Khan RA (2010) Application of Fuzzy Relations in Convalescing Link Structure. Bharati Vidyapeeth's Institute of Computer Applications and Management. 2(2):217–222

32. Gedeon, T.D. and Kóczy, L.T. Hierarchical co-occurence relations, IEEE International Conference on System Man and Cybernetics (SMC'98), session: Data Analysis & Information Science, San Diego, 2750-2755. (1998)

33. Gedeon, T.D., Wong, K.W., and Tikk, D. Constructing Hierarchical Fuzzy Rule Bases for Classification, The 10th IEEE International Conference on Fuzzy Systems, Melbourne, Australia, December 2001, pp. 1388-1391. (2001)

34. Gedeon TD, Kóczy LT, Wong KW, Liu P. Effective fuzzy systems for complex structured data. Proceedings of IASTED International Conference Control and Applications (CA 2001), Banff, Canada, June 2001, pp. 184-187. (2001)

35. Mendis BS, Gedeon TD. Aggregation selection for hierarchical fuzzy signatures: A comparison of hierarchical OWA and WRAO. IPMU'08. (2008)

36. Ballagi Á, Kóczy LT, Gedeon T. Robot Cooperation without Explicit Communication by Fuzzy Signatures and Decision Trees. In IFSA/EUSFLAT Conf. (pp. 1468-1473). (2009)

37. Gedeon T, Copeland L, Mendis BS (2012) Fuzzy Output Error. Australian Journal of Intelligent Information Processing Systems 13(2):37–43

38. Copeland L, Gedeon T, Mendis S. Fuzzy Output Error as the Performance Function for Training Artificial Neural Networks to Predict Reading Comprehension from Eye Gaze. In International Conference on Neural Information Processing (pp. 586-593). Springer International Publishing. (2014)

39. Wong, K.W., Fung, C.C., Ong, Y.S. and Gedeon, T.D. Reservoir characterization using support vector machines. In Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on (Vol. 2, pp. 354-359). IEEE. (2005)

40. Picek, S., McKay, R. I., Santana, R., & Gedeon, T. D. Fighting the symmetries: The structure of cryptographic boolean function spaces. In *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation* (pp. 457-464). ACM. (2015)

41. Sharma N, Gedeon TD, Mendis BS. Evolutionary algorithms using cluster patterns for timetabling. Intelligent Decision Technologies. 1; 7(2):137-50. (2013)

42. Rastogi R, Mondal P, Agarwal K, Gupta R, Jain S. GA Based Clustering of Mixed Data Type of Attributes (Numeric, Categorical, Ordinal, Binary and Ratio-Scaled). BVICAM's International Journal of Information Technology. 7(2). (2015)

43. Chande SV, Sinha M. Genetic algorithm: a versatile optimization tool. BVICAM's International Journal of Information Technology (BIJIT), Vol. 1, No. 1, page 7 to 12, year 2009.

44. Huang Y, Gedeon TD, Wong PM (1998) Spatial interpolation using fuzzy reasoning and genetic algorithms. Journal of Geographic Information and Decision Analysis 2(2):223–233

45. Saxena, AK, "On the Importance of Ensemble Classifiers" BVICAM's International Journal of Information Technology (BIJIT), Vol. 5 No. 1, page no. 569-576, year 2013.

46. Zhu D, Gedeon TD, Taylor K (2009) Keyboard before Head Tracking Depresses User Success in Remote Camera Control. Proceedings INTERACT 2009:319–331

47. Zhu, D., Gedeon, T.D. and Taylor, K. Exploring Camera Viewpoint Control Models for a Multi-Tasking Setting in Teleoperation, Proceedings of 29th ACM Conference on Human Factors in Computing Systems (CHI 2011), Vancouver, BC, Canada, pp.53-62. (2011)

48. Zhu D, Gedeon TD, Taylor K (2011) "Moving to the Centre": A gaze-driven remote camera control for teleoperation. Interacting with Computers (IwC) 23(1):85–95

49. Keerio MU (2009) Obstacle Avoidance through Visual Teleoperation. Bharati Vidyapeeth's Institute of Computer Applications and Management. 1(2):85

50. Asthana, A., Sanderson, C., Gedeon, T.D. and Goecke, R. "Learning-based Face Synthesis for Pose-Robust Recognition from Single Image," Proceedings of the British Machine Vision Conference BMVC2009, London, UK, 7-10 Sep. 2009, BMVA, pp. 1-10. (2009)

51. Gregor S, Lin A, Gedeon T, Riaz A, Zhu D (2014) Neuroscience and a nomological network for the understanding and assessment of emotions in information systems research. Journal of Management Information Systems (JMIS) 30(4):13–48

52. Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Gedeon, T., Breakspear, M. and Parker, G. A comparative study of different classifiers for detecting depression from spontaneous speech, In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference, pp. 8022-8026, IEEE. (2013)

53. Dhall A, Goecke R, Gedeon T (2015) Automatic group happiness intensity analysis. IEEE Transactions on Affective Computing 6(1):13–26

54. Singh, S., Gedeon, T.D. and Rho, Y. Enhancing Comprehension of Web Information for Users with Special Linguistic Needs, Journal of Communication, pp. 86-108. (1998)

55. Wong, K.W., Fung, C.C., Gedeon, T.D. and Chai, D., Intelligent Data Mining and Personalization for Customer Relationship Management, Proceedings of the Eighth International Conference on Control, Automation, Robotics and Vision, ICARCV 2004, Kunming, China, Dec 2004, pp. 1796-1801. (2004)

56. Sharma, N., & Gedeon, T. Hybrid genetic algorithms for stress recognition in reading. In European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics (pp. 117-128). Springer Berlin Heidelberg. (2013)

57. Copeland L, Gedeon T, Mendis S (2014) Predicting reading comprehension scores from eye movements using artificial neural networks and fuzzy output error. Artificial Intelligence Research 3(3):p35

58. Chow, C., & Gedeon, T. Classifying document categories based on physiological measures of analyst responses. In Cognitive Infocommunications (CogInfoCom), 2015 6th IEEE International Conference on (pp. 421-425). IEEE. (2015)

59. Kim J, Thomas P, Sankaranarayana R, Gedeon T, Yoon HJ (2015) Eye-tracking analysis of user behavior and performance in web search on large and small screens. Journal of the Association for Information Science and Technology 66(3):526–544

60. Al-Ismail, M., Gedeon, T., Sankaranarayana, R., & Yamin, M. Big 5 personality traits affect m-learning preferences in different contexts and cultures. In *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on* (pp. 1378-1382). IEEE. (2016)

61. Copeland, L. and Gedeon, T. Tutorials in eLearning; How Presentation Affects Outcomes. IEEE Transactions on Emerging Topics in Computing. 10.1109/TETC.2015.2499257 (2017)

62. Sharma, N. and Gedeon, T. Stress Classification for Gender Bias in Reading, B.-L. Lu, L. Zhang, and J. Kwok (Eds.): ICONIP 2011, Part III, LNCS 7064, pp. 348-355, 2011, Springer-Verlag Berlin Heidelberg (2011)

63. Li, X., & Gedeon, T. Gender disparity and the creepy hill in face replacement videos. In *Cognitive Infocommunications (CogInfoCom), 2015 6th IEEE International Conference on* (pp. 413-418). IEEE. (2015)

64. Sharma, N., Dhall, A., Gedeon, T. and Goecke, R. Modeling stress using thermal facial patterns: A spatio-temporal approach, In Affective Computing and Intelligent Interaction (ACII), Humaine Association Conference, pp. 387-392, IEEE (2013)

65. Irani, R., Nasrollahi, K., Dhall, A., Moeslund, T. B., & Gedeon, T. Thermal Super-Pixels for Bimodal Stress Recognition. In IEEE International Conference on Image Processing Theory, Tools and Applications. IEEE. (2016)

66. Vo T, Gedeon T. Reading your mind: EEG during reading task. In International Conference on Neural Information Processing (pp. 396-403). Springer Berlin Heidelberg. (2011)

67. Fei J, Pavlidis I (2010) Thermistor at a distance: unobtrusive measurement of breathing. IEEE Transactions on Biomedical Engineering. 57(4):988–998