Connectionism, Creativity and Guided Walks

Chris Thornton
Cognitive and Computing Sciences
University of Sussex
Brighton
BN1 9QH
UK

Email: Christopher.Thornton@firenet.uk.com
Tel: (44)1273 678856

May 21, 2003

1 Introduction

In the past decade or so Artificial Intelligence (AI) has seen the emergence of several new approaches. These include *geneticism*, which investigates the properties of genetic algorithms and the like, *connectionism*, which investigates the properties of networks of neuron-like units, and, most recently, *alife*, which focuses attention on some of the properties of living organisms. These approaches tend to focus on computational methods inspired by natural phenomena. Geneticism, for instance, investigates methods inspired by the structures and processes of evolution as described by the neo-classical theory. Connectionism, on the other hand, focuses on methods which are inspired by observations of basic neuronal activity in biological organisms. Finally, alife-ism focuses on computational aspects of phenomena such as self-replication, and investigates robotic methods inspired by insect motor-control mechanisms.

In addition to these and other relatively recent approaches, mainstream or 'classical' artificial intelligence continues to be an active area of research. Thus the AI researcher faces a field densely populated with alternative approaches. Some appreciate the advantages that such variety confers (1). But there are disadvantages to contend with too. There is, for example, what I call the 'other bandwaggons' problem. This is the tendency for a worker in one tradition to feel that the solution to any really difficult problem must be available more readily in another tradition. All too often, however, the other tradition serves up not a real solution but simply a neatly restyled reinvention of an old wheel.

Does this apply to connectionist models of creativity? Creativity is certainly a difficult problem in classical AI. And connectionism has certainly been perceived by some as offering the beginnings of a better explanation (2). But do we have any good reason for expecting connectionism to shed new light on creativity? And do those connectionist accounts which have been developed offer any real advance over classical accounts?

The starting point for any study of creativity is generally the observation that creativity involves the creation of something *new*. This observation is qualified with the observation that creation *ex nihilo* is a logical impossibility. (You can't create something out of nothing.) Therefore creativity must be the creation of something new out of things that already exist. But if this fact is logically necessary then it seems to imply that combination theories of creativity (i.e. theories which propose that creativity involves combining old stuff to make new stuff) must be essentially correct. (We may object to such theories, of course. And even if they are essentially correct they may not tell us very much. The statement that the internal combustion engine converts potential energy into kinetic energy is correct but not very informative.) Thus we really want to know whether connectionist accounts can or do take us beyond the basic premise of 'creativity as combination'.

In general, connectionist systems are constructed from units which, like biological neurons, store a certain level of activation, and links which, like biological synapses, propagate activation from one unit to another. There are three main types of system. Firstly, there are Hopfield-type networks (3). These nets, which typically feature total connectivity, tend to settle to a state in which unit activations are maximally compatible with connection weights. (This can be roughly characterized as 'positive weights between same-state units but negative weights between different-state units'.) If the network weights are loaded so as to capture constraints in a certain domain then the settling of the network performs a kind of constraint satisfaction. Of course, constraint satisfaction is a well defined computational process that can be implemented in many different ways. The Hopfield net is simply a particularly effective method for certain types of problem.

A second major category is the Kohonen-type network (4). These networks utilize a settling process similar to the one we see in Hopfield nets in order to perform a kind of constrained clustering (5). The aim of the process is to cluster the elements into an n-dimensional grid such that next-door neighbours are always maximally similar. These networks have generated a lot of interest partly because they provide an account of the way in which the mammalian visual system forms topographic maps. However, from the computational point of view they are most conveniently seen as a type of clustering procedure.

Last, but not least, we have Backpropagation-type networks (6). These networks are arranged so that activation flows forwards through several layers of units. There is a well-defined training procedure for weights which can be used to (try to) obtain desired activations of the output units for given activations of input units. The essence of this procedure is very straightforward. If a unit needs to have a higher level of activation than it is observed to have, the weights on its connections from active units should be increased. A complementary rule applies to the situation in which the unit needs to have a *lower* level of activation.

(The desired activations of non-output units are derived by seeing what impact we would like *their* activation levels to have on output units.)

Over the past ten years or so connectionism has attracted a great deal of attention but a careful study of the mechanisms involved reveals that, computationally speaking, it offers what are in effect 'new twists on old tricks'. In fact, it might not be too inaccurate to say that, for AI in general, connectionism can be thought of as an interesting and novel 'programming language' — one that is very good for certain tasks (such as capturing the 1st-order statistics of attribute databases). So what then should we make of connectionist models of creativity?

Dekker and Farrow's paper in this section is an interesting case. They describe an approach in which a Kohonen-type net is trained so that its weights define an n-dimensional space whose structure captures certain critical relationships of a given target domain. Using the trained net, they are able to obtain certain meaningful computational effects: the structure of the space represented by the network is such that transitions from 'problem' to 'solution' can sometimes be made solely on the basis of random (chaotic) stimulations.

This approach seems to get well away from classical combination theories but, on the other hand, it might be viewed as an implementation of a Markov chain (i.e. an implementation of a network of probabilistic state transitions). The system effectively traces a path through the transition network making appropriate random choices (given the fixed probabilities) at each choice point. Viewed from this perspective the model is suggesting that creativity is like a search or a 'guided walk'.

Guided walks also seem to play a central role in the scheme mapped out by Yao's paper. Yao considers the ways in which one might use simulated genetic processes to obtain useful architectures and learning rules for connectionist-style training. Here guided walking plays at least three roles. Firstly, there is the guided walk (conducted by the genetic process) which produces the design for the network architecture. Secondly, there is the guided walk which produces an appropriate learning rule. Finally, there is the guided walk (conducted by the connectionist procedure) which actually produces the solution. With models such as Yao's there is always the worry that the combinatorial complexity of carrying out several, related guided walks will be too great for current technology. However, the state of empirical testing has not reached the point at which such questions can be properly resolved.

In her major work on creativity, Boden (7) provides an extensive discussion of the way in which connectionist models can shed light on creative processes. She concentrates primarily on Hopfield-type 'settling' networks but also looks briefly at a model from the Backpropagation family (the 'past-tense' model of Rumelhart and McClelland (8)). She notes that Hopfield nets can reproduce certain properties which seem to be important within creative processes such as pattern-completion and graceful degradation. But here again, the waters are a little muddy since these properties might arguably be attributed to the level of representation used. Where a strongly 'micro' level of representation is used, 'macro' properties such as pattern completion and generalization are more easily obtained.

It is undeniable that the sorts of processes which seem to be inherent in connectionist models have the right 'feel' about them — they are fuzzy, error-prone, and somewhat unpredicatable (just like creative human beings) — but, on the other hand, we have to remember that even models based on the Hopfield network perform what in computational terms is no more than a guided walk. Technically speaking, Hopfield nets do nothing more than move continuously in the direction of the steepest energy gradient until they reach a local minimum. Everything else is just the context that *we* read in to the computation.

The situation then is still unclear. It seems to be generally agreed that a process that produces random combinations of existing entities cannot be regarded as creative. Or, putting it another way, primitive combination theories won't wash! But to be completely honest, it is not at all obvious that new connectionist models have really managed to get away from this very simple idea. Too often, rich system descriptions decompose all too readily into rather mundane computational processes — processes in which no-one would hope to find even the faintest glimmer of creativity.

As has been suggested, in the extreme case, connectionist models load full responsibility for the creative process onto the very basic and limited process of 'guided walking' or guided search (performed by, e.g., Hopfield nets, Kohonen nets, or hybrids). There may be very little to object to in the notion that creativity is, in some sense, a guided walk. However, it does beg the question: *what does the guiding?* If we look carefully, we find that in many models, the thing that does the guiding is the implementer of the system and/or the training set used. The buck is thus passed back but not actually 'grounded'.

This is one reason why Thornton, in his paper in this section, focusses attention on Hinton's family-tree example, in which a conventional backpropagation process is used to generate internal representations. In this example, the learning is not in any obvious sense guided towards the internal structures that it produces, and yet once those structures have been produced it is very easy to understand the function they serve within the task performed by the network. It is rather unfortunate that there are relatively few examples like this, in which simple connectionist learning procedures produce novel internal structures with a clear functional role. Other notable cases include Elman's work on language prediction (9) and, of course, the well-known work on the NETtalk system (10).

The somewhat impoverished state of creativity modelling is perhaps inevitable given our weak understanding of related processes such as memory, problem-solving and knowledge representation. Arguably, the attempt to model the full-blown creative process is premature. Our efforts might perhaps be better directed towards less ambitious goals. There is, for example, an active area of research which is primarily aimed at producing flexible support for the *human* creative process. (See the papers in the last section of this volume.) This work does not aim to model the process as such. Rather it attempts to provide 'materials' which allow the human creator to more easily develop and test out new ideas.

A domain in which this approach is receiving increasing attention (from both the academic and commercial communities) is that of music composition. The development of cheap but high-quality polyphonic synthesizers and the emergence of an international standard communication language for electronic musical instruments (MIDI) has given birth to a software development industry for 'composer environments'. These environments enable composers to put together pieces of music in a variety of ways. They typically provide an electronic but more flexible variant of the traditional pen->scrore->orchestra method for producing music. But increasingly they also provide higher-level tools for music creation. In sophisticated cases they may allow the composer to incrementally create new *languages* in which to express musical ideas and themes.

The structure of such composition languages can provide a fascinating insight into the 'upper strata' of the creative process of composition. By providing the composer with the infrastructure for reifying her more abstract musical knowledge, the computer effectively provides itself with the possibility of producing combinations and variations expressed purely in terms of the composer's *own* creative concepts. Of course, such software development work addresses the problem of creativity obliquely. It does not attempt to answer the big question 'what is creativity?'. But in helping to expose aspects of the process that would normally remain hidden, it can take us a few steps towards a far more detailed articulation of that question.

The time is ripe, then, for a certain degree of open-mindedness and eclecticism. Connectionist accounts of creativity are only just beginning to emerge. If, in certain cases, they seem to be reducible to partially discredited classical theories, we should not be unduly surprised. Software developments in the realm of electronic music composition may turn out — in the short run — to be a surprisingly fertile source of insights. At present, as Boden comments, creativity is still effectively a *mystery*, i.e., 'a question that can barely be intelligibly asked, never mind satisfactorily answered' (2: 1). To keep the momentum of progress we need to fully explore every new avenue of investigation.

References

[1] Torrance, S. (1992). Editorial. *AISB Quarterly*, No. 79 (p. 1).

[2] Boden, M. (1990). *The Creative Mind: Myths and Mechanisms.* London: Weidenfeld and Niicolson.

[3] Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, *79* (pp. 2554-8).

[4] Kohonen, T. (1984). *Self-organization and Associative Memory.* Berlin: Springer-Verlag.

[5] Everitt, B. (1974). *Cluster Analysis.* London: Heinemann.

[6] Rumelhart, D., Hinton, G. and Williams, R. (1985). Learning internal representations by error propagation. ICIS report, Institute for Cognitive Science, UCSD.

[7] Mooney, R., Shavlik, J., Towell, G. and Grove, A. (1990). An empirical comparison of symbolic and connectionist learning algorithms. In J.W. Shavlik and T.G. Dietterich (Eds.), *Readings In Machine Learning* (pp. 171-177). San Mateo, California: Morgan Kaufmann.

[8] Mozetic, I. (1986). Knowledge extraction through learning from examples. In T.M. Mitchell, J.G. Carbonell and R.S. Michalski (Eds.), *Machine Learning: A Guide to Current Research* (pp. 227-231). Boston: Kluwer Academic.

[9] Elman, J. (1990). Finding structure in time. *Cognitive Science, 14* (pp. 179-211).

[10] Sejnowski, T. and Rosenberg, C. (1987). Parallel networks that learn to pronounce english text. *Complex Systems, 1* (pp. 145-68).