ELSEVIER

Journal of Cognitive Systems Research 2 (2001) 235–240

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Book review

Review of The Mind within the Net: Models of Learning, Thinking, and Acting
Manfred Spitzer; MIT Press, Cambridge, MA, xi + 359 pp; ISBN 0-262-19406-6 (PB + HC)

Action editor: Stefan Wermter

Reza Farivar*

Department of Psychology, McGill University, 1205 Dr. Penfield Ave., Montreal, QC, Canada, H3A 1B1

Received 1 January 2001; accepted 1 February 2001

''How does the brain work? How do billions of neurons bring about ideas, sensations, emotions, and actions? Why do children play, and why do they learn faster than elderly people? What can go wrong in perception, thinking, learning, and acting?''

From this starting point, Manfred Spitzer begins his attempt to educate the wider public about the neural network revolution. In this effort, he reviews a diverse variety of neurophysiology and neurobiology, as well as experimental psychology data to demonstrate, as a minimum, the usefulness of neural network theory in understanding the nature of the human mind.

Who is this book meant for? In short, everyone: from the layperson to the specialist. The book attempts to bring together information from fields such as neurobiology, neural network theory, psychology, and psychiatry in order to better educate professionals such as physicians, psychologists, teachers, and other interested individuals.

This is no easy undertaking. To inform the general public of ''how the mind works'' is a dangerous thing: what if we are wrong about how the mind works? This is an important point to consider, given how many times the field of psychology has been wrong about this and how many people have suffered and perhaps continue to suffer as a result of our false conclusions. However, as one reads Spitzer's book, one realizes that by a basic understanding of neural network theory, one comes closer to the instructions to the mind. In fact, if you cannot wait to build up to this, you can quickly turn to the last page in the text to find Spitzer's ''User's Manual for Your Brain''. This manual would make much more sense, of course, if one has read and digested the contents of this excellent book.

The book is divided into three parts: Introduction, Principles, and Applications. The Introduction is specifically meant for the reader who is completely unfamiliar with neural network theory. The history of Neuron Theory is described, dating back to Camillio Golgi and Santiago Ramó y Cajal and the discovery of the neuron as the building block of the nervous system, Freud's rudimentary network models of psychological disorders, to the theorem of McCulloch and Pitts that neurons process information and not energy. The rest of the introductory chapter is written so as to arouse interest in the topic: very little is told about *how* and *what* networks are, but more focus is placed on the successes of the applications of neural network theory in understanding cognitive phenomenon. I was especially happy to see a good segment dedicated to describing how brains are different from computers, as it seems to

*Tel.: + 1-514-398-6150; fax: + 1-514-398-4896.
E-mail address: reza@ego.psych.mcgill.ca (R. Farivar).

1389-0417/01/$ – see front matter
PII: S1389-0417(01)00023-7

me that too many people, even in academia, are stuck on the computer metaphor of the mind.

The rest of the first part is dedicated to describing, lightly, some principles of neural network theory, including the idea of how a network is created, how learning occurs, and what could make connectionists believe that some sort of vector transformation takes place in our brain. At each point, Spitzer augments his excellent descriptions with vivid and clear examples, and so the reader is never left alone in understanding the challenging ideas and principles of connectionism.

Some of the examples are somewhat outdated. One is the past-tense acquisition model of Rumelhart and McClelland (1986). This first connectionist model of past-tense learning has received plenty of criticism, much of it valid, and newer, better network models of past tense learning have been formulated (e.g. Plunkett & Marchman, 1996). Although I praise Spitzer for keeping with tradition and reviewing some of the earliest accomplishments of the field, I think it is more appropriate to present the latest works that have corrected the mistakes of the past.

Spitzer addresses some crucial points in this chapter, such as ''Yes, neural networks seem interesting, but how can they actually *tell* us anything about the mind, given that they seem to be as complicated as the brain?'' The answer is simply that neural networks are *simulations* of cognitive processes, and as such, allow for all the advantages that simulations provide in other fields such as engineering, physics, meteorology, and so on. In the simulated networks, we can kill neurons, give them overdose of neurotransmitters and neuromodulators, change their connections, place restrictions, and even change completely how they learn; few people would be willing to undergo such experimentation, even if the ethics boards allow it. Appropriate warnings are given, however, that simply because a network appears to behave the same way as a human does not allow one to conclude that the two are working the same way; the network merely represents a possible mechanism by which a cognitive process may function. Also, as discussed much later in the book, a network model need not be biologically plausible to be functionally valid: the goal of a network is to capture the critical aspects of a phenomenon and to provide testable hypotheses for empirical studies.

In the introduction, Spitzer takes a very small shot at the Nurture/Nature debate: How do genes affect cognition? The author takes a quasi-mathematical attempt at this by noting that the human genome cannot contain enough information to code for all human behaviour, because behaviour is a result of connections between neurons, and those connections are too numerous. As a result, the brain cannot be pre-wired, but connections in the brain are learned. Although his discussion seems to be valid at this level of a discussion, it does not sufficiently describe the interaction of genes and learning. For one, genes *do* limit the possible architecture of the brain in that every normal human is born with the same lobes that function to process specific information; our eyes do no feed directly to our frontal lobes, for example. As such, genes must be determining the task that different areas of the brain must perform, and thus limiting the structure of the brain. These structural constraints do interact with learning, as neural networks demonstrate: an architecture ideal for one task may not suffice for another task. Furthermore, genes are likely to be involved in determining the distribution of neuromodulators and neurotransmitters in the brain, and these factors affect cognition very strongly. Thus, by simply counting the number of genes in the human genome and the number of connections in the brain, one cannot come to the conclusion that brains are created and formed purely by learning, but at least some interaction of genes and learning must be admitted. I would suppose that the author is trying to persuade the general reader to move away from believing too much in genes, and this is welcomed.

How does learning take place in the brain? This is an issue that the author tackles in the third chapter, describing long-term potentiation, the Hebb learning rule, and finally, the Generalized Delta rule, or backpropagation learning. The many diagrams and figures aid the reader to visualize synaptic transmissions with relative ease and thus enhance the reader's understanding of synaptic transmission and learning.

After the description of learning in the brain and learning in simulated neural networks, readers arrive at one of the most controversial issues in connectionist science: how can such a biologically implausible learning rule such as backpropagation be used to understand learning in the brain? It is

encouraging to see that Spitzer addresses such a question head on, not by simply citing papers that address the issue, but by providing an example where a backprop network has simulated very closely the functioning of the brain.

It is also interesting to see this issue raised in such an introductory book, but it goes to show that this is one of the thorniest issues that connectionists must contend with. Perhaps we should use more biologically plausible algorithms, but then we are running the risk of not playing with an idea before it is a proven fact. Hebbian learning, as biologically plausible as it may be, is not the most efficient form of learning and many problems cannot be learnt through the application of the Hebb rule. As a result, we toy with other algorithms, such as backprop, and we find that it simulates our experimental findings. But does the fact that backprop is not biologically plausible lessen its value in understanding human mental life? No, a simulation is meant to be an abstraction, and this abstraction need not become more refined unless empirical findings dictate an explanatory gap between the simulations and the data at hand.

This issue brings me to my one criticism of this book: that poor algorithms are heavily emphasized. Spitzer is on one hand attempting to persuade the reader away from the genetics-is-everything view of life and psychology, and on the other hand, he is describing static algorithms that would convince a reader that genes really must be everything! When one learns that in backpropagation networks, the simulator must choose the number of hidden units and all the parameter settings, one would feel that *these* settings must be genetic, since they are not present in the algorithm itself. And because the reader can learn that the number of hidden units present in a network can affect how well a network learns, the reader may be lead to believe that genes thus truly control the most important aspects of cognition and learning by limiting the number of neurons that are present in any given system.

Such is not a problem in some newer learning algorithms that generate hidden units as they are needed. One such algorithm is Cascade Correlation (Fahlman & Lebiere, 1990), which has been successfully used to model various developmental cognitive phenomena such as the balance scale (Shultz, Mareschal, & Schmidt, 1994), seriation (Mareschal & Shultz, 1999), and others. This algorithm generates feed-forward cascaded networks by beginning with a network of only input and output units and adding units as needed. An exemplary network may begin with only direct connections between input and output units, and these connections would be trained until error stagnates. When this happens, the direct connections between input and output units are frozen, and a pool of candidate units is generated, where each candidate unit receives connections from the input units and any hidden units, but is not connected to the output units. Training commences again on these weights, with the goal of maximizing the correlation between the activation of the candidate units and the presence of error in the networks. The candidate unit with the highest correlated activation to error is then recruited in the network, with its input side connections frozen while the output side weights are trained. This continues until error has been minimized.

In effect, Cascade-Correlation works by recruiting a unit that knows something about the error of the network (since the unit's activation correlates with the error) and then getting that unit to use what it knows about the error to minimize the network error (by training the output-side weights). The power of this algorithm lies in the fact that very few assumptions must then be made by the modeler, and thus some of the guesswork (which is referred to as ''artwork'' by the author) is taken out of the simulation. Furthermore, Cascade-correlation is more biologically plausible than backprop in that it works analogous to various neurological findings such as a burst in neural capacity (recruitment of a hidden unit), ossification of learning ability with time (as in the freezing of input site weights), burst in metabolic activity after neurogenesis and synapse formation (akin to the training of output-side weights), and synaptic pruning (discarding the unwanted candidate units after training) in a simple design (Shultz & Mareschal, 1997). Such an algorithm does not imply that genes control the parameter settings in the human brain, but that it is possible for some of these parameter settings to arise as a result of learning.

The discussion on learning implies something about how we encode information in our surroundings: if information presented to a network simula-

tion can be represented in a vector, and the networks final learned performance is the result of some complex vector transformation, then vectors must somehow exist in the head! Through the lucid and illuminating description of works by Georgopoulos and colleagues (1986, 1988, as cited in the book) on the representations of direction in the brain, Spitzer teaches the reader about the usefulness and the strength of representing knowledge as vectors instead of symbols. In the Georgopoulos experiments, rhesus monkeys were first trained on positioning a lever over several lights on a table when the lights lit up. These lights were equidistant, as they formed a circle, and the monkey's task was to first position the lever in the center of the circle (which caused one of the lights to turn on) and to then position the lever over that light. By making single-cell recordings, the researchers were able to observe that many neurons in the motor cortex of the monkeys were direction-dependent: they became active only for movement in certain directions. However, it was not the case the single neurons were active for a given direction, but that certain neurons fired within a range of directions.

Because vectors have both direction and magnitude, they were quite useful in representing the activity and the preferred direction of motor neuron in the monkeys' cortex. As a result, the researchers were able to quantify and present their results very precisely. This is a sharp contrast to the symbolic processing view that is common in popular psychology, where it is envisioned that our brains are like computers, with a hard-drive and RAM, where we store programs that somehow interpret our surroundings and other programs that act on those interpretations. Clearly, such a view is less than ideal in most circumstances, and clearly so in the case of the Georgopolous findings.

Part II of the book, termed ''Principles'', attempts to bring together the information provided in the first part in a more coherent and more widely applicable paradigm. These chapters are on higher-level topics, such as cortex maps, action of hidden layers, implications of neuroplasticity, and feedback in the brain.

Can cortical topography be simulated in a neural network? The answer is yes, with the application of Kohonen networks. These simple networks are composed of only an input and an output layer, with the two layers being fully interconnected (each input unit has output to every output unit). Furthermore, the units in the output layer are interconnected, where each unit activates its immediate neighboring units, while inhibiting the units further away, with the inhibition decreasing as the distance from the active unit is increased.

Learning in Kohonen networks is achieved through the application of the Hebbian rule, where connections between two co-active units are strengthened. What is of interest is that the same principles that affect cortical reorganization and topography affect the topographical formation in Kohonen networks: Topography is a function of similarity of inputs, their frequency, and their importance.

The chapter on hidden units is an elementary description of the function of a hidden unit, and why networks with hidden units can learn better than simple perceptrons or other network structures without hidden units. The notion of function approximation by neural networks is then briefly discussed in the context of a model of autism.

It is an excellent feature and a strong goal of this book to present psychological disorders in the connectionist framework. In the case of autism, it has been observed that some cortical and subcortical brain areas of autistic individuals have more neurons than in the same areas of an average person's brain. In a network model, these extra neurons can be thought of as extra hidden units that can help in remembering more information (which may explain why some autistic individuals have amazing memory capabilities), but may be inhibitive when these individuals are required to generalize from one situation to the next. As a result, autistic individuals have a decreased capability for abstract thinking; however, as Spitzer suggests, this deficit may have some fixes, and the network model guides this therapy: in order to increase generalization, one must focus on a more varied stimuli and thus forcing the brain to learn more abstract and higher-level recognition of the environment.

Why is it that, unlike a computer, we do not shut-down or crash when a part of our brain dies? How do we compensate for all the lost neurons that result from brain damage? By neuroplasticity, the ability of our brain to make up for lost connections due to neuron loss by forming new connections between the remaining neurons.

Again, some excellent examples are chosen in describing neuroplasticity to a wide audience. The introductory example is that of artificial ear implants, which present an organized set of inputs to the cochlea, but the organization of the inputs does not match that of the cochlea. However, people with such implants eventually learn to understand human speech as their auditory cortex adjusts to the new organization of the inputs. This reorganization brought about by neuroplasticity is one of the most evolutionarily advantageous features of our nervous system, as it allows us to recover some lost functions after brain damage. A model of phantom limb phenomenon is also discussed in relation to Kohonen networks to further illuminate the concept of neuroplasticity.

Of course, Spitzer tries to leave nothing out, and thus describes recurrent networks in a chapter on feedback in the brain. Hopfield networks are described briefly, but more emphasis is placed on Elman's (1991) recurrent network model. At first, recurrence is described in the context of working memory for a network. The most interesting example is that of grammar learning in children and its simulation using recurrent neural networks. The Elman (1991) model was, in its primary form, fed inputs of complex and simple sentences, but could not extract any rules from this. However, when it was given simple sentences, it could learn some basic rules, and by then increasing the complexity of the sentences, the network was able to finally learn even more complex rules. Yet, infants are not only exposed to simple sentence structures, and so something else must be simplifying the inputs that they receive if the model is to be correct. What was found with further simulations was that by reducing the number of context units (units which receive input from other hidden units and feed this input *back* into the hidden units, thus maintaining a form of working memory) the network was unable to learn complex rules from complex sentences, but it did extract *simple* rules from those complex sentences, as if the complex sentences where actually simple ones; by slowly increasing the number of context units, the network was eventually able to extract more complex rules, even though it was given complex sentences from the start.

The last part in the book, titled ''Applications'', is really what Spitzer wants to talk about, but could not do so without some explanations before. The first chapter of this part deals with knowledge representations in neural networks, drawing the conclusion that information need not be represented by symbols, but can be represented in terms of connection weights. This is particularly illuminating when the discussion of Piaget is brought into the picture, where stage development (generally viewed as a static, rule-governed global behaviour of the growing brain) is described in terms of connectionist models. One example used is the balance-scale phenomenon. It has been demonstrated that children at different ages perform quite differently on this task, which involves judging the tilt direction of a balance, given different weights on each side of the scale, at different distances from the fulcrum. Four stages have been identified in this task:

> *Stage 1:* Children only take the size of the weight into account.
> *Stage 2:* Only distance of the weights from the fulcrum is taken into account only when weights are equal on both sides.
> *Stage 3:* Both weight and distance are taken into account, but inconsistently.
> *Stage 4:* Consistent answers are given, incorporating both weight sizes and distances from the fulcrum.

Again, in criticism of this book, Spitzer sites an older model of this phenomenon, that of McClelland (1989), which, as Shultz et al. (1994) point out, is limited in that (a) it required many limiting assumptions (proportionally greater equal-distant training examples, forced weight separation between weight and distance information in connections of the hidden units), and (b) the network never clearly reached Stage 4. However, a later model by Shultz et al. (1994) using the Cascade-Correlation generative learning algorithm did generate proper stage development (with a clear development of Stage 4 without necessitating any forced weight changes). Again, it would have been nicer if more recent models were discussed, that better represent the successes of connectionism.

The chapter on semantic networks is mainly a precursor to the later presentation of models of schizophrenia. However, this chapter is, on its own, filled with good examples and a special emphasis is

placed on category specific naming deficits and how they can be modeled using categories developed by Kohonen networks. The results of such modeling efforts are compared to PET and fMRI studies on brain activations during naming, showing that cerebral blood flow in certain regions of the brain do correlate strongly with the presentation of members of a specific category, suggesting that the brain may be doing something like a Kohonen network when learning new words and categories.

It is not unlike a psychiatrist to be very interested in schizophrenia, and thus, Spitzer dedicates a full chapter to a discussion of this mental disorder, presenting information from the fields of brain imaging, electroencephalography, and, of course, neural network modeling.

Schizophrenics have odd semantic networks, says Spitzer. In comparison to normals, who normally activate only related terms in a semantic priming study, schizophrenics tend to generate abstract mappings, and thus show a more disordered semantic priming effect. But why? Based on the semantic network model, by increasing the size of the spread of activation, one may activate more distant semantic contents, and as a result, demonstrate a disordered semantic priming effect.

This is further implicated in the analysis of speech pauses by thought-disordered schizophrenic patients during a description of a scene: whereas normal patients made pauses when switching between words that were within the context of the scene and words that were out of context, thought-disordered schizophrenic patients did not make such a pause, demonstrating that both context and non-context semantic information may be activated in such patients. This is in concordance with the semantic network model of the disorder, in that a larger spread of activation would predict a larger amount of activation of non-context words in comparison to normal individuals.

The results from modeling are then discussed in the context of the neuromodulator, dopamine, which the author implicates in controlling the spread of activation in our brains. This chapter on schizophrenia is too large, diverse, and filled with too many excellent and illuminating examples to be covered in a review. I can only say that Spitzer has succeeded in presenting connectionism as a new paradigm that has the potential to incorporate many psychological phenomena, including mental disorders. In fact, the book is seasoned with numerous examples of mental disorders, including autism, Alzheimer's disease, and schizophrenia, among others. This book is thus an excellent entry point for the mental health professional who wishes to learn more about formal modeling using neural networks.

The concluding chapter pulls together all the presented information and describes how, in general, human mental life can be better understood in terms of neural activity, as modeled in artificial neural networks. By now, the reader is in good shape to read the ''User's Manual for Your Brain'', some of which confirms our intuitions and popular notions about learning, development, and mental health, but within the context of the formal models presented in the book, carries the greater weight of a strong argument deserving our avid attention.

References

Elman, J. L. (1991). Learning and development in neural networks: the importance of starting small. *Cognition 48*(1), 71–99.

Fahlman, S. E., & Lebiere, C. (1990). The cascade-correlation learning architecture. In: Touretzky, D. S. (Ed.), *Advances in neural information processing systems*, vol. 2, Los Altos: Morgan Kaufmann, pp. 524–532.

Mareschal, D., & Shultz, T. R. (1999). Development of children's seriation: a connectionist approach. *Connection Science 11*(2), 149–186.

McClelland, J. L. (1989). Parallel distributed processing: Implications for cognition and development. In: Morris, R. G. M. (Ed.), *Parallel distributed processing: implications for Psychology and Neurobiology*, Oxford: Clarendon Press, pp. 9–45.

Plunkett, K., & Marchman, V. (1996). Learning from a connectionist model of the acquisition of the English past tense. *Cognition 61*, 299–308.

Rumelhart, D., & McClelland, J. L., (1986). On learning the past tense of English verbs. In: Rumelhart, D., & McClelland, J. L., (Eds.), *Parallel distributed processing: explorations in the microstructure of cognition*, vol. 2, Cambridge: MIT Press, pp. 216–271.

Shultz, T. R., & Mareschal, D. (1997). Rethinking innateness, learning, and constructivism: connectionist perspectives on development. *Cognitive Development 12*, 563–586.

Shultz, T. R., Mareschal, D., & Schmidt, W. C. (1994). Modeling cognitive development on balance scale phenomena. *Machine Learning 16*, 57–86.