On Association Techniques in Neural Representation Schemes

John A. Barnden
Computer Science Department
Indiana University, Bloomington, Indiana.

Section 1: Introduction

It has often been proposed that, in the brain, associations between information items take the form of suitable settings for synaptic weights [e.g. Anderson and Mozer (1981), Anderson et al (1977), Fahlman (1979, 1981), Feldman (1981), Feldman and Ballard (1982), Goddard (1980), Hebb (1949), Hinton (1981), Kohonen et al (1981), Wickelgren (1979)]. An information item is implemented as a potential or actual pattern of neural activity in some particular set of neurons. An association from item I to item J is implemented as the existence of suitable synaptic weights on neural paths from the neuron set for I to that for J, such that the active presence of the I pattern tends to cause the J pattern to appear. The patterns are anchored, in the sense that the identity of the particular neurons whose activity constitutes a pattern is crucial. In what I shall call the *dedication* approach [Feldman (1981), Feldman and Ballard (1982), Fahlman (1979, 1981), Goddard (1980), Hebb (1949), Wickelgren (1979)], all or many of the neurons in the neuron set for an information item are individually dedicated to that item, in that they do not appear in the neuron sets for other items. (The dedicated neurons are often called "grandmother", "pontifical" or "cardinal" cells.) In what I shall call the *sharing* approach [Anderson and Mozer (1981), Anderson et al (1977), Hinton (1981), Kohonen et al (1981)], individual neurons in the set generally belong to the sets for many other items as well.

We shall look at various problems facing currently proposed schemes which encode association by means of synaptic weight values, when they try to account for ***rapid, complex information-processing*** such as is involved in understanding and producing natural-language or acting in the world. Some of the schemes do address certain specialized types of short-term processing, but as far as I am aware they do not deal in any general way with the problems to be discussed. We have space here for no more than a brief look at the problems. A more detailed paper on the subject is in preparation.

Section 2: Some Problems for Synaptic-Weight Schemes

We adopt the working hypothesis that we must show how the neural mass could act as an implementation of semantic network processing of the sort typically postulated in AI and cognitive psychology. We shall assume networks in which relationships as well as non-relational items are coded as nodes, the links being left for restricted "syntactic" uses such as linking a relationship node to the nodes for the partakers in the relationship. We take the nodes to be implemented as neural activity patterns, and we take the links to be implemented as individual associations encoded as synaptic weight settings. These assumptions are in reasonable accord with the approaches taken in Fahlman (1979, 1981), Feldman (1981), Feldman and Ballard (1982), Hebb (1949), Hinton (1981), Kohonen et al (1981), and Wickelgren (1979). The problems on which we focus are association (link) deletion and node marking.

Network Alteration and Traversal

If semantic network manipulations occur in short-term cognition, presumably they include the traversal of networks and the modification of networks (whether by a change of graph structure or by replacement of one node by another). Such an alteration presumably involves replacement of the piece of net to be altered by a new piece of net which is linked in properly either (a) to the rest of the net as it is or (b) to a copy of the rest of the net. The replaced piece in case (a), or all of the old net in case (b), must somehow be put out of play, whether by being marked, inhibited in some way, or isolated by having associations leading into it deleted. Explicit deletion of an association (link) has received little attention in work on synaptic-weight schemes, and it raises difficulties. Presumably the neural substrate must be put into something like the condition it would have been in had the association not been present. This is no great problem in a dedication scheme, because the synaptic weights defining the association can be reduced to some small value. In a sharing scheme, however, the synaptic weights are important also to other associations, so the particular state they are left in depends crucially on what those associations are. The trouble is that, unless the whole network is somehow traversed or rebuilt, the deletion operation has no guidance as to what those associations are. Similar comments apply to the idea of "inhibiting" a piece of net. By this I mean leaving the structure essentially intact but switching on some special agency which stops the nodes in question from being activated, and/or which stops the associations from being used. The problems in the remaining alternative, marking, receive attention below.

To turn to traversal of networks, it is standard for traversal algorithms implemented on computers to use a marking scheme to ensure that parts of the structure are not traversed more than once. A form of marking would for the same reason be needed in a neural implementation of traversal. To avoid marking we could suggest a random-walk process, which may well duplicate parts of the traversal. This would appear to be a rather error-prone and/or time-consuming technique, and its adoption in a neural model would require convincing argumentation.

Marking

Fahlman's (1979, 1981) dedication scheme makes heavy use of marking to effect certain types of inference and structure matching. Hinton (1981) uses a marking scheme for structure matching which is more sophisticated but similar in spirit. Marking can be used to get the effect of pointers. If we have some way of activating nodes which are marked in a specific way, then we get the effect of following a pointer.

How is a node marked? We note first that we should be able to provide marking of several different types, and that nodes must be able to be unmarked as well as marked. The first suggestion is that the nature of the potential firing of the individual neurons in the node's neuron-set is changed. For instance, the firing trains of a neuron could have several distinct possible patterns, corresponding to different types of marking. The objection to this is that it involves a major change in the philosophy of synaptic-weight schemes, where it is usually assumed that neurons can be more or less active (e.g. can fire at higher or lower frequency) to indicate the "confidence" with which the node is present, but cannot be active in symbolically distinct ways. Once the door is opened to distinguishing between modes of firing for symbolic purposes, the question arises of whether significant symbolic information of more general sorts should not be encodable in firing patterns. Also, in a sharing scheme, a neuron contributes to several nodes, some of which may be marked and some not.

A second suggestion is that the node to be "marked" is replaced temporarily by a new node which acts as a marked version of the original node. But here we are appealing to an operation of structure alteration as discussed above. As we saw, the replaced node must be put out of play. In a sharing scheme, it seems that the most viable alternative is to put it out of play by marking it! We therefore have a vicious circle. In a dedication scheme, the replaced

node could be inhibited or isolated, but this is a cumbersome process if done merely for the purpose of marking, especially when we consider the possible need for later unmarking of the node.

A third possibility is to have a special node which acts as an explicit mark and is put into association with the node to be marked. Even in the dedication scheme this is an over-elaborate proposal. At one extreme, we have the possibility that there is just one special node, so that all the neural sets implementing nodes which might ever need to be marked have to be connected by a neural path to this special mark node. At the other extreme, we could have a distinct special mark node for every distinct non-mark node. At either extreme, or anywhere in between, a large amount of 'hardware' is set aside just for marking purposes. A sharing scheme faces analogous difficulties, but also faces a particular difficulty in unmarking: the mark, or at least its association to the marked node, must be put out of play, but in the present case we do not want to mark a mark! We might get round this problem by stipulating that the neurons used in the mark node cannot be used for anything except marking, i.e. that they are dedicated to marking; perhaps then the association between mark node and marked node can be easily broken (in view of our comment above that deletion of an association does not appear to be a problem in a dedication scheme). The methodological objection to this is that we are diluting the purity of the philosophy of sharing schemes by letting in dedication in restricted cases for ad hoc reasons.

The last suggestion we make is that extra neurons are somehow included temporarily in the neuron set for the node. We could suppose that each dedicated neuron set has several subsets of special neurons, one subset per marking type. A node is considered to be marked if and only if the appropriate subset of neurons is activated when the main neurons for the node are activated. However, this marking technique is rather ad hoc, since it requires the idea of marking to be built into the very hardware of the system, and we begin to wonder why we should not allow specialized neuron sets to hold symbolic structures more complicated than marks. The sharing schemes face a further difficulty. It is not at all clear how the process of unmarking would work unless the special mark neurons for a node are distinct from the neurons used by any other node; but then we have a restricted form of dedication much as in the third proposal.

A general observation about all the above marking proposals is that they treat marks as data items which need special mechanisms for their implementation or use. This contrasts with marking in computers, where an algorithm may use data items in such a way that we call them marks, but where those data items require no special mechanisms for their implementation or use.

Section 3: An Alternative

It is of interest that an alternative which avoids the problems of Section 2 can be provided. We shall continue to assume that the issue is the implementation of semantic networks and their manipulation in the course of short-term information-processing. We suppose still that nodes are implemented as patterns of neural activity. However, our patterns are substantially ***unanchored*** with respect to neural location: their precise positions in the neural mass are irrelevant. To make sense of this, we suppose that the neural areas in which the patterns reside are regular in structure. In fact, we make the tentative, simplifying assumption that the areas are physically structured as arrays (typically of dimension two). We call each such area a ***pattern matrix*** (PM). A PM is an array of neural circuits called PM elements. Each PM element can be active in any one of a certain number (say a dozen or two) modes, any combination of modes being allowed. Short-term information structures are patterns of PM activity over the PMs. The pattern in a PM is generally composed of well-defined subpatterns, some of which play the rôle of network nodes. The precise position of a subpattern is unimportant, although its position relative to other subpatterns may be crucial, as we shall see in a moment.

A distinctive feature of our scheme is the way in which subpatterns can be in association so as to form structures. The association does not take the form of transmission-facilitated neural paths. Instead, association is by ***adjacency*** and ***similarity***. Adjacency association is similar to the association of data items in a computer by virtue of their being in adjacent locations. So, two subpatterns in a PM which are adjacent to each other may be taken to be associated. (As a special case, one of the subpatterns may be like a closed boundary and contain the other subpattern.) Similarity association is akin to content addressing in computers. There is a mechanism attached to PMs such that the presence of a subpattern in a PM can cause sufficiently similar subpatterns in this and other PMs to be be "highlighted" by high activity in some mode (whose identity is passed to the mechanism as a parameter). Subpatterns which are thus associated by similarity can be placed adjacent to other subpatterns (e.g. nodes) which they can be considered to "label". The labelled subpatterns can thereby be regarded as being indirectly associated by a combination of adjacency and similarity association. A further form of association is derived from adjacency association: two subpatterns in different parts of a PM can be joined by a line-like supattern (whose ends are adjacent to the first two subpatterns). Such line-like subpatterns are analogous to the link lines in a diagram of a semantic network. They are also analogous to the neural paths in synaptic-weight schemes, but instead of a facilitated transmission path there is a path of activated neural networks (PM elements).

We suppose that connected to the PMs there is a neural mechanism embodying a production system. The condition part of a rule responds to the presence of fairly simple combinations of primitive subpatterns. The primitive subpatterns in the case of network-like patterns would be subpatterns acting as nodes, links and labels (or perhaps components of labels). The action part of a rule is able to insert subpatterns, delete subpatterns, copy subpatterns, move subpatterns around in PMs, follow link-like subpatterns, highlight subpatterns, invoke the pattern-similarity association mechanism and communicate with mechanisms outside the PM production system. The set of rules does not change in the short term. We envisage the rules to be implemented as neural networks attached to the PMs, and the triggering of rules to occur by just the sort of neural associative techniques as are proposed in synaptic weight schemes. Also, patterns in PMs could associate to long-term structures by such mechanisms. It should be noted that subpatterns in PMs are not themselves the full embodiments of concepts or other entities; rather they are merely symbols for or "ambassadors" of those entities.

In our scheme all information in a semantic network, including association, is encoded as activity patterns. This provides greater uniformity and elegance than is present in synaptic-weight schemes, in which there are two competely different embodiments of information: (potential) patterns of activity and synaptic weights. At the same time, by being relatively close to the way computers work (if we take adjacency, similarity and line-linking to correspond to locational adjacency, content addressing and direct addressing in computers) our proposal has the advantage that techniques developed for information-processing in computers can relatively easily and plausibly be supposed to occur in the brain. For instance, nodes and links can be labelled and marked very simply and naturally, either by highlighting them or by placing label subpatterns next to them. No special extra mechanisms need be postulated, and our deeming a particular feature of a PM pattern to be a mark is merely a result of the way particular rules use the feature (cf. the comments on marking in computers at the end of Section 2). Subpatterns which are associated by adjacency, line-linking or label-similarity can be altered without affecting their inter-association, because the patterns embodying the associations are independent of the patterns associated. Associations can easily be deleted, whether by moving subpatterns so that they are no longer adjacent or by removing labels or line links. (Such removal is performed by suppressing the activity in the PM elements concerned.) Therefore, the problems of deletion and marking that we noted in Section 2 do not arise for our scheme, so that operations such as traversal no longer present special difficulties.

There is no space here to present the scheme in more detail. One version of the scheme is reported in Barnden (1982a, 1982b). Barnden (1982b) goes into considerable detail concerning the possible operation of the production system in manipulating network-like patterns in PMs. The scheme appears to be no less well supported by known facts about the brain than are

the synaptic-weight schemes – indeed, by virtue of its foundation on arrays of neural nets it fits in more naturally with the regular and ubiquitous columnar organization of cortex [Mountcastle (1978)] than those schemes do.

References

Anderson, J.A. and Mozer, M.C. Categorization and selective neurons. In Hinton and Anderson (1981).

Anderson, J.A., Silverstein, J.W., Ritz, S.A. and Jones R.S. Distinctive features, categorical perception, and probability learning: some applications of a neural model. *Psychological Review,* 1977, *84, 413-451.*

Barnden, J.A. The integrated implementation of imaginal and propositional data structures in the brain. *Procs. 4th. Annual Conf. of the Cognitive Science Society,* Ann Arbor, Michigan, 1982a.

Barnden, J.A. A continuum of diagrammatic data structures in human cognition. Tech. Rep. 131, Computer Science Dept., Indiana University, October 1982b.

Fahlman, S.E. *NETL: A system for representing and using real-world knowledge.* Cambridge, Mass.: MIT Press, 1979.

Fahlman, S.E. Representing implicit knowledge. In Hinton and Anderson (1981).

Feldman, J.A. A connectionist model of visual memory. In Hinton and Anderson (1981).

Feldman, J.A. and Ballard, D.H. Connectionist models and their properties. *Cognitive Science,* 1982, *6,* 205-254.

Goddard, G.V. Component properties of the memory machine: Hebb revisited. In P.W. Jusczyk and R.M. Klein (Eds.), *The nature of thought: Essays in honor of D.O. Hebb.* Hillsdale, N.J.: Lawrence Erlbaum, 1980.

Hebb, D.O. *Organization of behaviour.* New York: Wiley, 1949.

Hinton, G.E. Implementing semantic networks in parallel hardware. In Hinton and Anderson (1981).

Hinton, G.E. and Anderson, J.A. (eds) *Parallel models of associative memory.* Hillsdale, NJ: Lawrence Erlbaum, 1981.

Kohonen, T., Oja, E. and Lehtiö, P. Storage and processing of information in distributed associative memory systems. In Hinton and Anderson (1981).

Mountcastle, V.B. An organizing principle for cerebral function: the unit module and the distributed system. In G.M. Edelman and V.B. Mountcastle, *The mindful brain.* Cambridge, Mass.: MIT Press, 1978.

Wickelgren, W.A. Chunking and consolidation. *Psychological Review,* 1979, *86,* 44-60.