VIRTUAL UTERAL INHIBITION IN PARALLEL ACTIVATION MODELS OF ASSOCUTIVE MEMORY*

James A. Reggia

Department of Computer Science
University of Maryland
College Park, MD 20742 USA

ABSTRACT

This paper describes a new theory of how spreading activation may occur in associative memory models formulated as parallel activation networks. The theory postulates that competition for activation by nodes/concepts in a network is a fundamental principle of memory retrieval. Using only excitatory connections between concepts, a specific implementation of this model is able to demonstrate "virtual lateral inhibition" between competitors and other interesting behaviors that have required use of explicit inhibitory connections in the past.

I INTRODUCTION

During the last several years there has been a great deal of interest in AI in determining what kinds of parallel architectures best meet the needs of various AI tasks. This paper is concerned with "value-passing systems," networks in which the processing elements communicate by passing around continuous quantities (numbers) and by performing simple arithmetic operations on these values [4]. Such architectures are often intended as models of associative memory, and frequently they are characterized by an analogy with neurobiological networks and processing paradigms. Recent examples include "connectionist models" [5], "interactive activation models" [10], the Boltzman machine [7] and ACT ML

This paper presents a new "competition-based" theory about how spreading activation may occur in value-passing associative memory models. First some terminology and the need for a model of spreading activation that can support "virtual lateral inhibition" are discussed. Then a theory is introduced which postulates that competition between cognitive activities for limited resources is a fundamental organizing principle of memory retrieval. A specific instantiation of the theory is used to illustrate the concepts involved.

•Supported by the Office of Naval Research, the National Science Foundation (DCR-8451430), and Software Architecture and Engineering, Inc.

II PARALLEL ACTIVATION MODELS OF MEMORY

In memory models implemented as value-passing systems, each processing *node* typically represents a "concept" or "hypothesis", and the *level of activation* associated with a node represents the relevance of or confidence in the concept/hypothesis represented by that node [1,5,73. For this reason, and because of the distributed nature of the computations involved, the term *parallel activation network* is used for such models. Nodes communicate with each other using links which have one or more weights associated with them. If the link from node A to node B has a positive weight, indicating that activation of node A tends to increase activation of node B, then such a link is *excitatory.* Conversely, a negative weight indicates an *inhibitory* connection. There are both excitatory and inhibitory interconnections at the level of neuronal circuits in the nervous system. It is therefore not surprising that both types of interconnections have been adopted by way of analogy in higher-level cognitive networks.

It has often proven convenient to view parallel activation networks as partitioned into *layers* of conceptually similar nodes, and to conceive of information as initially entering one "lowest" layer (e.g., features) and propagating "upwards" to others (e.g., letters) [2,10], When nodes in a lower layer directly inhibit nodes in a higher layer such inhibition may be called *forward inhibition* since it is in the direction of initial flow of activation. Conversely, inhibition in the opposite direction may be called *backward inhibition,* and inhibition between nodes in the same layer can be referred to as *lateral inhibition.* Lateral inhibition has long been recognized to play an important role in contrast enhancement in neural networks and has been argued to be an important aspect of selective behavior [6]. Explicit lateral inhibitory links have also often been used in parallel activation models of associative memory in cognitive science to produce selective behavior [2,5,10], representing one way in which these recent cognltlvely-oriented models have borrowed processing paradigms from earlier neural modeling studies. For example, Figure 1 shows two mutually Inhibitory nodes n_1 and n_2. With a typical model for spread of activation even a transiently higher input to one node (say n_1) followed by equal inputs to both nodes can lead to stabilization of activation with one node maximally activated (n_1 and the other having zero activation (n_2). This "winner-take-all" phenomenon [5] comes about because of the lateral inhibitory connections.

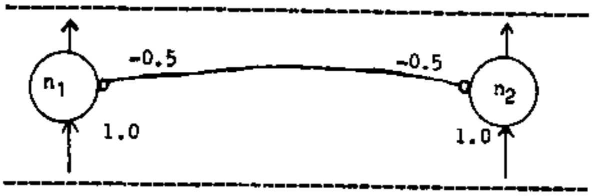

Figure 1: Lateral inhibition using explicit inhibitory links between two "competing" nodes (modified from [5]). Weights are indicated adjacent to connections.

While <u>lateral inhibition</u> is a useful concept in parallel activation models, the use of <u>explicit lateral inhibitory links</u> to achieve lateral inhibition poses a number of difficulties for implementing real world models of associative memory. At the cognitive level, in contrast to the neurobiological level, there is at best limited empirical evidence for such *explicit* "negative associations" between the concepts involved. Published tabulations of the associations between related concepts usually include only positive weights (frequencies). Since these empirical studies do not provide negative weights between competing concepts, one is faced with the problem of determining how such weights are to be assigned in building parallel activation networks where weights are based on probabilitites [11], Furthermore, as the number of "competing" nodes grows, the number of mutual lateral inhibitory connections needed can grow quite rapidly. For example, in the print-to-word mapping, a single character node may connect to thousands of word nodes, each of which would apparently require a mutually inhibitory connection with <u>all</u> of its competing word nodes [10],

In the following an approach to parallel spreading activation is proposed that can resolve these problems. Specifically, this approach is capable of producing *virtual lateral inhibition:* apparent lateral inhibition between competing nodes in the absence of explicit lateral inhibitory connections. This phenomenon is produced without giving up the notion that computations in parallel activation networks should be local: each node can only "see" the activation and link weights of its immediately adjacent neighbor nodes.

III <u>COMPETITION-BASED SPREAD OF ACTIVATION</u>

In previous parallel activation models, the activation flowing into a node <u>is not determined by the level of activity possessed by the receiving node</u>. Typically, the incoming activation "seen" by a node n_i is a weighted sum or some other function of the activities of its neighbors [1,2,5,6,7,10], This incoming activity is used by node n_i to update its own activation level, which is then distributed by node n_i during the next increment of time as output to appropriate neighbors.

In the competition-based parallel activation theory described here, the spread of activation is determined in a different way. When a node n_i assumes an activation level above its normal resting level, its neighbors <u>actively</u> compete for the "energy" possessed by node n_i. Further, the ability of a neighboring node to compete for n_i's activity or resources <u>is proportional to that neighbors existing level of activation</u>. Resources acquired by a neighbor node in this fashion occur at the expense of resources that are available to its competitors, leading one to predict that virtual lateral inhibition will occur. The metaphor used here is that the "stronger" a node is (i.e., the higher its level of activation), the more effectively it can compete with other nodes for a source of energy/activation.

To examine this idea, one possible instantiation of competition-based spreading activation is presented below (other formulations are possible and are currently being explored). In the formulation given here, the presence of an underlying associative network is assumed where the nodes in the network are partitioned into disjoint layers as illustrated in Figure 2. For convenience, we also assume that each node in one layer is directly connected to at least *one* node in each adjacent layer. (These assumptions are <u>not</u> an essential part of the theory.)

Each connection between a node in one layer and that in another has two weights reflecting the directionally-oriented frequencies of association between the "concepts" represented by the nodes. For example, if n_i is a node in one layer associated with a node n_j in a different layer, then a bidirectional link appears between n_i and nj in the network. One weight "attached" to this link is w_{ij} representing the frequency (conditional probability estimate) with which the concept represented by n_i occurs given that the concept represented by n_j is known to be present. Similarly, the attached weight W_{ji} represents the frequency with which n_j occurs given that n_i is known to be present. We restrict such weights to $0.0 < w_{st} < 1.0$; the non-negative nature of these weights distinguishes them from the possibly negative "synaptic weights" in neural models and in psychologically-oriented parallel activation models of associative memory [5,103. Nodes in this example network do not explicitly inhibit other nodes in either the same layer or in adjacent layers. Furthermore, permitting weights to be highly asymmetric (wij W_{ji} in general) also distinguishes this approach from others [7,8].

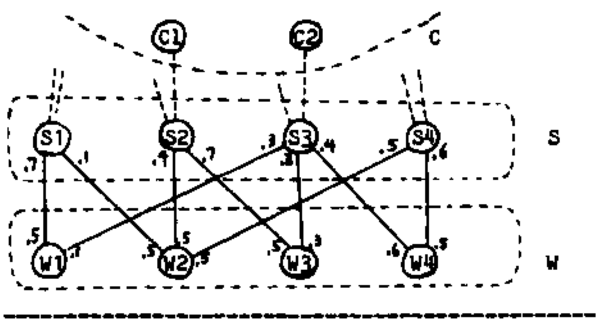

Figure 2: A word-sense disambiguation network.

As a specific "toy" example, consider a grossly simplified version of the problem of mapping a word into the appropriate word sense (Figure 2). Layer W represents word/morpheme units (assume there are just four possible words), layer S represents possible word senses or "meanings" (again, just four), and layer C, which we ignore for the time being, represents a number of possible semantic contexts in which word sense disambiguation can occur. Note that each word node Wi relates to multiple senses, e.g., node W2 is connected to S1, S2 and S4, reflecting the ambiguity of words in natural language. Thus {S1,S2,S4} represents the nodes in layer S which are competitors for the title of "sense indicated by W2." Similarly, each word sense is connected to multiple possible words (e.g., synonyms).

In Figure 2, the weight W_{ji} corresponding to spread of activation from node n_i to node nj is written adjacent to n_j. The weights on "outputs" from a node in one layer to nodes in an adjacent layer sum to 1.0, consistent with our earlier definition of weights as conditional probability estimates. Each node n_i in memory has an activation level $a_i(t)$ at time t, constrained so that $0.0 \leq a_i(t) \leq 1.0$, and we let $a_i(t)$ represent the belief in the entity represented by node n_i. Each node also has a natural resting activation level r_i and a decay rate d_i reflecting how quickly $a_i(t)$ returns to the resting level in the absence of external influences. We make two assumptions. First, we assume that at time t the maximum rate at which node n1 is capable of distributing activation to its competing neighbors in an adjacent layer is proportional to $a_i(t)$. Second, we assume that n_i parsimoniously "desires" to support at most a total of one unit of activation in associated competing nodes in an adjacent layer. Thus, the total amount of activation n_i distributes to an adjacent layer at any point in time decreases as the collective activity of its neighbors in that layer rises. For example, in going from words to word senses, this second assumption states that each word maps onto exactly one of its possible senses. These two assumptions can be approximated by stating that the total output of node n^ to its neighbors at time t is given by

$$out_i(t) = (1.0 - \text{sum of activities of neighbors}) \cdot a_i(t)$$

as long as this quantity is non-negative, and zero otherwise.

The idea of competition is introduced into this model by permitting the neighbors of n_i in an adjacent layer to actively compete for the total output activity of n_i The ability of neighbor node n_j to compete for n_i's output activity $out_i(t)$ is proportional to its "strength" $a_j(t)$, and to the weight of association W_{ji}. Thus, the activation out $_{ji}(t)$ transferred from n_i to n_j at time t is given by

$$out_{ji}(t) * \left(w_{ji} \cdot a_j(t)\right) \cdot out_i(t)$$

(the symbol "*" is read "is proportional to"). "Stronger" neighbors of n_i therefore extract a greater portion of n_j's finite activation energy, leaving a smaller portion for "weaker" competitors. It is the fact that aj(t), the activation of the receiving neighbor node, appears in the formula for $out_{ji}(t)$ above that makes this a competition-based model. Finally we define the total flow of activity into a node n_j at time t to be

$$in_j(t) = \sum_i out_{ji}(t),$$

the sum of all its inputs from neighbors.

In summary, a specific form of spreading activation has been described where the portion of a node n_i's output activation going to neighbor node n_j is proportional to the ability of n_j to compete for that activity (reflected by the formula for $out_{ji}(t)$ above). Given this competition-based approach to distribution of activation, we can adopt an approach similar to that used by others to update a node's activation level (e.g., [10]). Let the symbol I|(t) indicate the net flow of activation into nn at time t, i.e.,

$$I_i(t) = in_i(t) - out_i(t).$$

Then the rate $a_i(t)$ at which n_i's activation changes is given by the net flow of activation into the node minus the decay:

$$\dot{a}_i(t) = I_i(t) \cdot \left(1 - a_i(t)\right) - d_i \left(a_i(t) - r_i\right).$$

The term $\left(1 - a_i(t)\right)$ in this latter equation assures that $a_i(t)$ approaches its maximum value asymptotically. Note that all of the above computations are local in the sense that n_i only needs to "see" the activations and weights associated with its immediate neighbors.

Two numerical examples are now presented to demonstrate the behavior of the above competition-based parallel activation model. These examples are based on the network in Figure 2, and were Implemented using PAN, a parallel activation network simulator. PAN is a LISP program that permits one to specify a network and method for spread of activation, and to describe external inputs to the network that are to occur during simulation, PAN then performs the indicated simulation while periodically displaying the activity level of nodes (PAN is similar in spirit to ISCON

[12]). In the examples, PAN is used with the specific competition-based model of spreading activation described above. We have run larger simulations (over 60 nodes) with different networks but space limitations prevent their discussion.

```
[NODE   W1  0.0  0.0)
(NODE   W2  0.0  0.0)
[NODE   W3  0.0  0.0)
(NODE   W4  0.0  0.0)
(NODE   S1  0.0  0.0)
(NODE   S2  0.0  0.0)
[NODE   S3  0.0  o.o)
 NODE   S4  0.0  0.0)
 PARIN  W1  (SI   0.5)   (S3  0.1))
 PARIN  W2  (S1   0.5)   (S2  0.5)  (S4  0.5))
 PARIN  W3  (S2   0.5)   (S3  0.3)1
 PARIN  W4  (S3   0.6)   (S4  0.5))
 SONIN  SI  (WI   0.7)   (W2  0.1)1
[SONIN  S2  (W2   0.1*)  (W3  0.7)]
 SONIN  S3  (WI   0.3)   (W3  0.3)  (W4  0.4))
 SONIN  S4  (W2   0.5)   (W4  0.6))
 INPUT  W3  0.1   6.0    1.0)
[DELTA  0.1 )
```

Figure 3: A simulation specification given to PAN simulator based on network in Figure 2.

Figure 3 illustrates a "simulation specification" as it is given to PAN. The first eight lines specify the four word and four word sense nodes of Figure 2, indicating that each node has both a resting activity and a decay rate of 0.0. The next four lines specify the PARent connections INto the word nodes along with appropriate weights, and the subsequent four lines specify the corresponding SON connections INto the word senses (compare with Figure 2). The next line, indicating external INPUT to the network, has the format

(INPUT node start-time stop-time amount).

Figure 3 thus indicates that an external INPUT of 1.0 units of activation enters word W3 during the first 6 units of time, simulating the occurrence of word W3. Next a constant DELTA of 0.1 units of time is specified, indicating the fineness of time quantization, and PAN is told to run the simulation for 20 units of time.

Upon giving the simulation specification in Figure 3 to PAN, the following output was produced by the above competition-based model:

Time	Hi	S2	S3	S3/S2
0	0	0	0	1.0
0.2	0.18	0.007	0.003	0.43
1	0.55	0.20	0.01	0.05
2	0.77	0.43	0.02	0.05
4	0.95	0.68	0.02	0.03
6	0.99	0.79	0.03	0.04
10	0.98	0.87	0.04	0.04
15	0.98	0.90	0.04	0.04
20	0.98	0.91	0.04	0.04

This simple "lossless" network illustrates a number of important properties possible with competition-based spread of activation. First, the spread of activation is circumscribed in that it radiates only to nodes S2 and S3, the senses evoked by W3. Activation levels in all other nodes remain at zero. Second, although the input stimulus terminates after six units of time, activity in the activated nodes subsequently remains stable. This is reminiscent of "stable coalitions[11] [5]. Third, this example demonstrates virtual lateral inhibition between S2 and S3, with the dominant S2 rapidly suppressing activity in S3 <u>indirectly</u> by absorbing the majority of activition available from W3. Past models of spreading activation have only produced such lateral inhibition by having explicit inhibitory connections between nodes like S2 and S3. Finally, note that there is a brief initial period of time when the ratio of activity S3/S2 is relatively high. This initial "window of opportunity" makes possible some interesting context effects, as illustrated below.

For the second example, we illustrate how a prexisting "context" can result in the appearance of a switch being thrown to redirect the flow of activation. To do this, we use the same simulation specification as that given in Figure 3, except the previous single external INPUT statement is replaced with the following two statements:

```
(INPUT   W3   0.1   6.0   1.0)
(INPUT   S3   0.1   6.0   0.2)
```

The first input stimulus is exactly the same as in the preceding example. The second weaker input (0.2) is directed to S3 (recall that S3 was indirectly inhibited by S2 in the previous example). This latter "input" is used for illustrative purposes to simulate positive feedback from the "context" C2 in which disambiguation of word W3 is occurring (see layer C in Figure 2). In this situation, PAN produces the following listing of activations:

Time	W3	S2	S3
0	0	0	0
0.2	0.18	0	0.05
1	0.57	0	0.32
4	0.96	0	0.81
10	0.99	0	0.93
15	0.99	0	0.95
20	0.99	0	0.95

In this example, the absolute contextual effect produced by retrograde excitation to node S3 "switches" or "gates" the flow of activity from node W3 completely to S3 (contrast with the first example above). The absolute switching here is contingent upon the resting level of *S2* and S3 being zero Initially; in networks where r_i might be a very small positive number, perhaps reflecting the relative frequency with which the concept represented by node n_i occurs, switching/gating of the sort demonstrated here directs a small amount of activation to S2.

IV DISCUSSION

This paper has presented an approach to spreading activation in parallel associative networks that is distinguished from previous approaches by its competition-oriented nature. A theory of competition-based parallel activation as a model of associative memory has not been significantly studied in AI in the past, although there is some relevant related work. For example, the idea of nodes as active agents bears some resemblance to "actors" [14 3, and the "contract net framework" involves at least an implicit notion of competition [133. However, both of these and similar models are "message passing systems" at the AI symbol processing level [4 3. Others have postulated competition and/or parsimonious allocation of activation/energy as important influences in cognition, but this work has been at the "hardware level" of neural modeling and has been formulated quite differently [3,7,8,9 3.

As illustrated above, even in the absence of explicit inhibitory connections between nodes, a competition-based approach to spreading activation can exhibit a number of important properties: virtual lateral inhibition between appropriate nodes, circumscribed activation, stability of activation ("stable coalitions"), and context effects (e.g., switching). Our research group is currently investigating the feasibility of developing a full-scale model of the cognitive activities involved in the real world print-to-sound mapping [11]. This task should provide an excellent test of the theory of competition-based spread of activation proposed in this paper.

REFERENCES

[1] Anderson J: The Architecture of Cognition. Harvard, 1983.
[2 3 Cottrell G: Model of Lexical Access of Ambiguous Words, Proc Nat Conf on AI, 1984, 61.
[3] Edelman G: Group Selection as the Basis for Higher Brain Function, in The Organization of the Cerebral Cortex. F. Schmitt et al, MIT Press, 1981, 535.
[4 3 Fahlman S: Three Flavors of Parallelism, Proc. 4th Nat. Conf. Canadian Soc. Comp. Studies Intell., Saskatoon, May 1982.
[5 3 Feldman J and Ballard 0: Connectionist Models and Their Properties, Cognitive Science, 6, 1982, 205-254.
[63 Grossberg S: Associative and Competitive Principles of Learning and Development, in Competition and Cooperation in Neural Nets. S Amri and M Arbib, Springer-Verlag, 1982, 295.
[73 Hinton G, Sejnowski T and Ackley D: Boltzmann Machines: Constraint Satisfaction Networks that Learn, TR-CMU-CS-84-119, Dept. of Comp. Sci., CMU, 1984.
[83 Hopfleld J: Neural Networks and Physical Systems with Emergent Collective Computational Abilities, Proc Nat Acad Science. 79, 1982, 2554.
[93 Kosslyn S: Personal Communication, 1983.
[103 McClelland J & Rumelhart D: An Interactive Activation Model of Context Effects in Letter Perception. Psych. Rev., 88, 1981, 375-407.
[113 Reggia J and Berndt R: Modelling Reading Aloud and Acquired Dyslexia, 1985, submitted.
[123 Small S, Shastri L, et al: ISCON: A Network Construction Aid and Simulator for Connectionist Models, TR-109, Dept. of Comp. Sci., Univ. of Rochester, 1983.
[133 Smith R: A Framework for Distributed Problem Solving, Proc. 6th IJCAI, 1979, 836.
[14] Yonezawa A and Hewitt C: Modelling Distributed Systems, Proc. 5th IJCAI, 1977, 307.