Marker-passing and Microfeatures

James A. Hendler

Computer Science Dept./Systems Research Center
University of Maryland, College Park

Abstract:
The cognitive modeling community is presently divided between two different approaches to the spread of activation through networks. One school holds that symbolic information must be progagated, the other that numeric weights are used and activation spreads in a more analog manner. In this paper we describe a mechanism which allows the two processes to be merged via the introduction into the symbolic network of a *defining-characteristic* link which is affects the spread of the symbolic information in a manner resembling local connectionist computations. We demonstrate that the combined system is more powerful than either of the separate models alone.

Introduction

Spreading activation, in the form of computer models and cognitive theories, has recently been undergoing a resurgence of interest in the cognitive science and AI communities. Two competing schools of thought have been forming. One technique, that of *marker-pasting,* is based on the works of Quillian [18] and concentrates on the passing of symbolic information through an associative knowledge representation. A second group, holding to the *parallel distributed proeessing[20]* or *connectionist* approach, has focused on the passage of numeric information through an associative network.

The primary advantage of the former group is that such systems gracefully interact with traditional AI symbolic processing models for natural language processing [4](II](15][I7] and planning [12) and accounting for the psychological results in various experiments (for example [6][2l][23)(24)). The connectionist approach, however, has proven useful for modeling various types of cognitive performance that the symbolic approach has traditionally had problems with: similarity based reasoning, learning, and categorization tasks. Both local connectionist models, such as those of [10],[22], and distributed models [14) seem to best account for results from the psychological literature on categorisation (discussed in [1][16]). Both groups have proposed massively parallel models for these computations. This is a corner stone of the connectionist approach, but various models massively parallel algorithms for performing symbolic marker-passer have been proposed [2] [9] [12).

It is our strong contention that neither a purely connectionist scheme nor a marker-passing approach which ignores connectionism is as powerful as a mechanism which combines some of the best features of each. This paper compares the connectionist and marker-passing approaches and shows why neither alone seems adequate, describes an implemented mechanism that solves provides integration between these approaches and handles some of those problems neither approach is sufficient to solve, and concludes by describing where this work is heading.[1]

The need for hybridisation

To justify the necessity of combining the competing formulations of activation spreading, we start by examining some specific strengths and weaknesses of each approach.

The marker-passing approach works in an associative network in which links representing symbolic information (rather than simply link strengths or the like) are present. Information is passed through these links to to find connections between nodes in a "mechanistic" way, as opposed to the approach of using inferencing algorithms consisting of a set of deductive rules. Marker-passing algorithms have been formulated which can pass very simple information [9] or complex marks which include various amounts of control information for use in the marker-passing process [4] [12] [18] .[2]

That the marker-pssing approach integrates well with standard AI systems can be seen, for example, in the word sense disambiguation system of Hirst [15] which model approaches ambiguity by setting up a set of frames that put word senses in one to one correspondence with semantic meanings. Each sense of a given word "competes" for recognition based on two criteria: the syntactic information available and the connectedness of the words to each other. This connectedness was found via a marker-passing system.

Hirst's system would handle the word "ball" in a phrase like *The rubber ball* by using marker-passing to find connections between the various word senses of the words in the phrase. When this phrase was encountered the system would pass markers between rubber and ball and a path such as

RUBBER—►(Isa rubber material) —► material
—►(composed-of physobj material) —►♦ physobj
—► (Isa sphere physobj) —► sphere —►
(word-sense ball sphere) —► BALL

would be found. This would cause the preference of the sphere meaning of ball over the dance meaning.

By integrating his word sense disambiguator into a syntactic parser Hirst was able to use syntactic clues in performing disambiguation. For example, the word "TIRE" in the sentence *The left front tire is flat,* can only be a noun, not a verb, and Hirst's system could use this information to ignore the VERB senses.

An alternate approach to word sense disambiguation is that of Cottrell [7] which uses a local connectionist system to perform the disambiguation. The correct interpretation of the sentence would be found at the end of a period of numeric computation by examining the nodes remaining activated in a stable configuration.

For Cottrell to produce a parse to provide syntactic information to the disambtguator would require the design of a syntactic parsing scheme within the massively parallel framework. This may not be impossible, in fact Cottrell [8] discusses some steps in exactly such a direction, but at present connectionist schemes are not having tremendous success at producing such results. In fact, a general weakness of connectionist approaches is the generation of sequential behavior which seems to be an integral part of language processing, planning, etc.

Marker-passing algorithms, too, have limitations, the primary of which is the need for explicit symbolic knowledge connecting the objects being examined. Consider how a system like Hirst's might handle the following two sentences:

John attacked Bill with the knife
John attacked Biii with the letter opener

1- Thii paper if a brief description of work described in more detail in fid). Implementation details and further justifications can be found therein.

The marker-passer would find the path:

ATTACK
—► Instrument of a physical attack is a weapon
—► weapon —► a knife is a weapon
—► KNIFE

and thus be able to recognize the first sentence as a physical attack (as opposed to say a heart attack or an attack by satire). Such a path won't be found for the second sentence unless we had specifically encoded the information that a letter-opener can be used as a weapon. Assuming such knowledge was not available, a system such as Hirst's could not use the marker-passer and would need to default to other, more complex mechanisms.

Finding similarities between objects has been a traditional problem for symbolic AI systems in much the way that sequential behavior has been a traditional problem for connectionist systems. While some knowledge representation schemes have a limited ability to do classification, this classification is generally viewed as a separate control process[1]. What we need in this case is a decomposition of an object into components and the categorization of a set of components during the marker-passing phase. Information must be provided to allow a marker to flow from letter-opener to knife. We could use a predicate such as *(IS-LIKE LETTER-OPENER KNIFE)* to solve this problem, but such a solution is clearly lacking in elegance and, in fact, will not solve the more general case .

What is needed here appears to be some sort of breakdown of the letter-opener into a set of features and a recognition that those features share many similar characteristics with a knife.

Recognizing similar characteristics, a major problem for symbolic systems, is one of the strengths of a connectionist model. A *microfeaturc* based analysis, which decomposes a single object into a set of smaller, usually perceptual, features is a technique which has been shown viable in the connectionist framework [14][16]. Activation strengths flowing from an object to a set of microfeatures, or from a set of microfeatures to an object are modeled via numeric weights in a network. Placing activation on one object can cause other similar objects to gain activation. Further, the use of microfeatures appears to be consistent with the results of psychological experimentation (as cited in [16]).

A hybrid mechanism

In this section we describe a mechanism that combines a marker-passing algorithm with a local connectionist, microfeature-based, network. The scheme we propose here consists of two parts: an extension to the knowledge representation, allowing the presence of the microfeaturc information, and an extension to the marker-passer which allows this knowledge to be used. The particular formulation of marker-passing we describe come from a planning mechanism called SCRAPS wherein the marker-passer improves the planning behavior in the presence of external information. The details of the marker-passer used in SCRAPS and a discussion of the cognitive implications of the work can be found in [12].

To extend the knowledge representation we need to recognize that our letter-opener is comprised of a set of features such as "pointed, metallic, and thin" and that a knife would have these as defining characteristics. It will not be enough, however, to perform a simple match on all these characteristics. The letter-opener will also have features which are not shared by the knife. We must move towards the connectionist notion of using an ensemble of microfeatures and an activation level.

Encoding the information about the sets of microfeatures which define a given object is fairly easy in our present framework. We define a set of microfeatures and a special network link called a *defining characteristic.* Thus, each of the objects in our network is linked to a set of microfeatures which "define" it (Figure 1).

To handle the extension of the marker-passer we need to create a situation in which an object's being marked by the symbolic system corresponds to the same object being "activated" in the connectionist manner. This activation then spreads through the distributed memory causing corresponding activations on the other objects in the network. These activations must then be returned to symbolic marks. In such a system marking LETTER-OPENER would cause a large amount of activation on KNIFE with smaller activations on GUN and ROPE. Adding some sort of threshold-like mechanism which would supress marker-passing on the less activated nodes would complete the system.

The implementation of such a mechanism requires blending two significantly different types of information. The symbolic marker-passer is passing discrete symbolic information that needs to propogate through the network, the distributed memory needs an activation strength that spreads via numerical combination. To integrate the two a mechanism must be used which can provide such a numerical activation strength to the distributed memory. It turns out, however, that a mechanism already existing for other purposes in our system can be used for this purpose.

One of the most important features in the design of a marker-passer is a mechanism which can limit the number of nodes marked and paths returned. Several such attenuation mechanisms have been proposed, but our marker-passer uses a mechanism which is designed to limit both length and branchout of the paths found without violating the locality constraints needed for massively parallel computing. This is done by using a numerical constant, called *zorch,* to start marker-passing and dividing it by outbranching as we proceed. If we start at a highly branching node, it and all its first set of descendants can be marked, but the zorch runs out quickly. If, however, we hit such a promiscuous node late in processing, its descendants are not marked since we do not have enough zorch left .

The hybrid system works by using the zorch reaching a node as activation energy over the network defined by defining-characteristic links. These links are used as a local connectionist network so that activation spreads in an "analog" manner over this subnetwork. As other nodes of this network gain activation they start passing marks with a zorch based on their activation. Marking then proceeds as normal, paths are found, and the path evaluator is invoked as normal.

In our example, letter-opener, appearing in the sentence, would originate marking with an initial zorch. This zorch is then divided among the branches leading out from the node. Some of this zorch goes through symbolic links to the properties of letter-opener causing the marking of the neighbors of this node (for example mail). The rest of the activation is used as the constant starting the flow of activation through the microfeature network. Those nodes sharing microfeatures gain activation proportional to the number of microfeatures shared. The node knife, sharing multiple

1- For example Brachman's [1] KL-ONE system provides a "classifier" which can be run to determine what class is the best fit for some particular configuration of objects. (An example of this use of Brachman's classifier is the language parser described in (Brachman and Schmolze. 1984)).

2- Consider the case of recognizing that some random "long, thin, pointed metal object," (call it OBJECT-27) could not be taken onto the plane. It would seem clearly wrong to assert (IS-LIKE OBJECT-27 KNIFE) without some sort of analysis of the object in terms of its features.

It turns out that designing the attenuation mechanism in this way has many desirable features[12].

microfeatures with letter-opener, receives most of the activation. This enables marking to continue from knife with a zorch proportional to the amount of energy it received from letter-opener. Marker passing continues and we find the path:

ATTACK
—▶ Instrument of a physical attack Is a weapon
—▶ weapon —▶ a knife is a weapon
—▶ knife —▶ shares activation with —▶
—▶ letter-OPENER

which is what we were seeking.

Although the informal description presented so far serves to describe the mechanism, there are still some details we've omitted. In particular, we have glossed over details of how symbolic information is carried through the distributed network.

To this point we have been describing marks as very simple entities. In reality, each marker in our system carries information about the origin of marking and the path that led to the node being marked, as well as information about the zorch which must be carried along with the activation energy that traverses the network. In our implementation, as the "strength" of the activation is computed throughout the distributed network, the information that originated the particular marking is carried. This has two major advantages, it allows several sets of activation to proceed through the network at the same time, and it allows the marker-passer to continue as activation at some node is collected (i.e. activation strength is computed without disturbing the symbolic information on the mark).

Allowing the distributed memory to be computing several activations simultaneously is an important feature of our mechanism. One of the goals of our formulation of marker-passing is to let each set of marking happen in parallel. In the example we've been discussing, marks are started from each word more or less As these marks flow through the network, they may encounter the microfeature network at any time. When we mark BILL and LETTER-OPENER some entities will gain activation from one, some from the other, and some from both. We must be able to sort out which is which, and our system uses the symbolics information as "tags" for this purpose. When a node is activated from several microfeatures its activation for each separate tag is computed.

Comparison to Connectionist models

The defining characteristic links described in this paper function as a form of distributed memory. There are still, however, many differences between this and a standard connectionist system, even a local connectionist system. Primary among these is that our network does not provide weighted links or inhibitory links. Both of these are generally found in connectionist models.

It is possible to add such items to our network — as the zorch is turned into activation strength and passed through the network the computations can be as arbitrarily complex as desired. If weighted links were used the zorch could be divided based on the weights. Inhibition, if desired, can be implemented by adding inhibition links between separate network nodes, or it can be simulated by using asymmetric weights and a competition based activation scheme [20].

We are currently working on a model which partially combines the competition-based scheme and our own. If a network, made up of both semantic information and the sorts of distributed information found in our defining characteristic nodes activates all and only the nodes corresponding to paths found by the marker-passer, then it becomes possible to integrate the path finding and evaluation functions more directly into the network. We believe that this may be closer to a neurobiologically plausible model of what is happening than the simple model described in this paper.

In Conclusion

In this paper we have demonstrated that a hybrid system allowing both symbolic and numeric spreading activation has advantages over either approach alone sine the strengths and weaknesses of the two approaches are complementary. Connectionist models are stronger at similarity based reasoning (as in finding the commonalities between the knife and the letter-opener), but they are particularly weak when it comes to providing control and symbol processing (as, for example, in performing planning). Models have been proposed for performing symbolic manipulations on massively parallel, distributed memories [25], but even these models have still needed to use a traditional VonNeuman style architecture to control gating and control flow within the system.

Since each style of machine has proven suitable for certain problems, our approach, demonstrated herein, has been to look for a blending. Essentially, the technique described in this paper took a massively parallel algorithm, the "most connectionist" of many of the modern symbolic approaches, and combined it with a local connectionist model, the most "symbolic" of the connectionist approaches. The "buying the letter-opener" example, presented in this paper, is an example of a problem in which merging these approaches enables us to deal with a task which neither approach, alone, has yet been able to solve.

We also believe this approach, the hybridization of symbolic and subsymbolic systems, provides an exciting area of exploration. Experiments have shown cognitive evidence for both connectionist and non-connectionist systems. Integration of these approaches is an important step towards explaining these results. We believe that the work presented in this paper is a start in the correct direction and an indication that progress can be made.

References

(1) Anderson, JR *The Architecture of Cognition,* Harvard University Press, Massachusetts, 1983

(2) Blelloch, Q.E. *AFL-1: A programming language for massively concurrent computers* MIT AJ Laboratory, Technical Report AI-TR 918 (forthcoming, 1986)

(3) Brachman, R J *A Structural Paradigm For Representing Knowledg* TR # 3605, Bolt, Baranek, And Newman, Ma May, 1978

(4) Charniak, EC "Passing markers A theory of contextual influence in language comprehension " *Cognitive Science* 7(5) July - Sept 1983

(5) Collins AM and Loftus, EF A spreading-activation theory of semantic processing *Psychological Review,* 6, 1972, *407-4S8.*

(6) Collins, AM and Quillian, MR Does category size effect categorization time? *Journal of Verbal Learning and Verbal Behavior,* 9, 1970, *482-488.*

(7) Cottrell, GW *A Connectionist Approach to Word Sense Disambiguation* Doctoral Dissertation, Computer Science Department, University of Rochester, May, 1985

(8) Cottrell, GW Connectionist parsing *Proceedings of the Seventh Annual Conference of the Cognitive Science Society,* 1985, *201-2*

(9) Fahlman, S E *NETL: A system for representing and using real-world knowledge* MIT Press, 1979

(10) Feldman, J A and Ballard, DH Connectionist models and their properties *Cognitive Science,* 6, 1982 *205-254.*

(11) Granger, RH , Eiselt, KP and Holbrook, JK *Parsing with Parallelism: A Spreading-Activation Model of Inference Processi During Text Understanding* Artificial Intelligence Project, University of California, Irvine Technical Report #228, Sept 1984

(12) Hendler, J A *Integrating Marker-passing and problem-solving: A spreading activation approach to improved choice in planning* i Doctoral Dissertation, Brown University, January, 1986 ii. Ablex Publishing Corp , NJ 1987 (in press)

(13) Hendler, J.A. Extending MaiM i-Passing to Microfeaturea: A cautious step towards the subsymbolic in Advances in Cognitive Science Vol. //N Sharkey (ed.) Ablex, N J (in press)

(14) Hinton, OE., McClelland, J.L., and Rum«lhart,D.E. Distributed Representations, in Rumelhart, D.E. and McClelland, J.L (eds.) ParaUel Distributed Computing MIT Press, Cambridge, Ma., 1986

(15) Hirst G J Semantic Interpretation Against Ambiguity Brown University Computer Science Technical Report CS-83-25, Dec 1983

(16) McClelland, J L , Rumelhart,D E and Hinton, G E The Appeal of Parallel Distributed Computing, in Rumelhart, D.E. and McClelland, J.L (eds.) ParaUel Distributed Computing MIT Press, Cambridge, Ma., 1986

(17) Norvig, P Recovering Intended Inferences from Text, Doctoral Dissertation, Computer Science Department, University of California at Berkeley, (forthcoming)

(18) Quillian, MR. Semantic Memory, Doctoral Dissertation, Carnegie Institute of Technology (Carnegie-Mellon University), 1966. Published as Report 2, Project 8668, Bolt, Baranek, and Newman Inc , 1966

(24) Tanenhaus, M.K., Lei man, J.M., Seidenberg, M.S. Evidence for multiple stages in the processing of ambiguous words in syntactic contexts Journal of Verbal Learning and Varbal Behavior, 18(4), 1976, 681-689.

(25) Touretcky, D.S. and Hinton, G.E. Symbols among the neurons: Details of a connectionist inference architecture Proceedings of the Ninth International Joint Conference on Artificial Intelligence, 1985, S38-243.

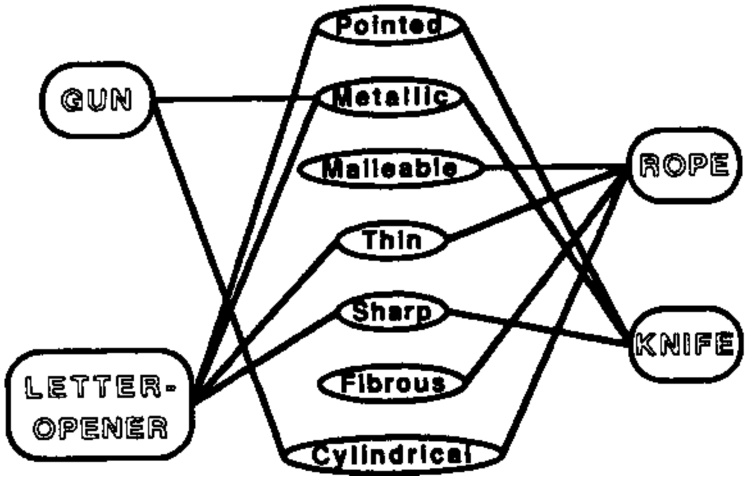

Figure 1: Defining-Characteristic links in a network

(19) Reggia, J. Virtual inhibition in parallel activation models of associative memory Proceedings of the 9th International Joint Conference on Artificial Intelligence, 1986, 244-148.

(20) Rumelhart, D.E and McClelland, J.L (eds) ParaUel Distributed Computing MIT Press, Cambridge, Ma, 1986

(21) Seidenberg, MS The time course of information activation and utilisation in visual word recognition Reading Research: Advances in Theory and Practice, Vol. 5 (ed.) Bosner, Waller, and MacKinnon, Academic Press, New York, 1985, 199-252

(22) Shastri, L. Evidential Reasoning in Semantic Networks: A formal theory and its parallel implementation Doctoral Dissertation, Computer Science Department, University of Rochester, Sept., 1985.

(23) Swinney, DA Lexical access during sentence comprehension: (Reconsideration of context effects Journal of Verbal Learning and Verbal Behavior, 18, 1979, 645-659.