Evidential Inference in Activation Networks

Jerome A. Feldman and Lokendra Shastri
Computer Science Department
University of Rochester

Introduction

Psychological and biological results suggest that many cognitive tasks like visual recognition, categorization and associative retrieval do not take more than 100 computational steps. This follows because typical neuronal firing rates are a few milliseconds and the response time of cognitive agents during numerous experimental tasks is a few hundred milliseconds. Given that most cognitive tasks require access to a large body of information, the above observation imposes a major constraint on the manner in which conceptual information may be organized and accessed by cognitive processes. In particular it seems to preclude an interpreter that examines the knowledge base. This paper briefly outlines a framework for organizing and accessing conceptual information that appears to offer several advantages over previous work [Fahlman 79]. The proposed framework suggests an evidential semantics for knowledge and describes how the above may be encoded as an active and massively parallel (connectionist) network [Feldman & Ballard 82]. The resulting system has been run on simple examples and is capable of supporting existing semantic network applications dealing with problems of recognition and recall in an uniform manner. The framework also provides a natural way of representing "inconsistent" or conflicting information and using it in making inferences. It embodies an important class of inference that may be characterized as working with a set of competing hypothesis, gathering evidence for each hypothesis and selecting the best among these. A detailed treatment of this framework appears in [Shastri & Feldman 84].

Overview

In the proposed framework, conceptual knowledge is organized as a network of active elements which interact with one another via controlled spreading of activation. The information encoded in the "memory" network is accessed via other network fragments, each of which is a connectionist encoding of a *routine*. We present a simple example to introduce the notation and the overall framework. Figure 1 depicts the interaction between a fragment of an agent's restaurant routine and a part of his memory network. The routine fragment decides whether some food goes well with red wine on the basis of the food's taste. A routine is represented as a sequence of nodes (units) connected so that activation can serve to sequence through the routine. In the course of their execution, routines pose queries to the memory network by activating relevant nodes of the memory network. The memory network returns the answer by activating appropriate units in the routine. We depict action steps as oval-shaped nodes, queries as hexagonal nodes and answer nodes as circular nodes. In this routine fragment, the task of deciding on a wine results in a query to the memory network about the taste of food and the decision is made on the basis of the answer returned by the memory network. Answer nodes in a routine mutually inhibit each other and the answer node receiving the maximum activation from the memory network triggers the appropriate action. The memory network in the example encodes the following information:

HAM and YAM are two concepts in the domain.
Concepts in the example domain are characterized by two properties, *HAS-TASTE* and *HAS-FOOD-KIND.*
HAM is SALTY in taste and is a kind of MEAT, YAM is SWEET in taste and is a kind of VEGETABLE.

Each arc in the network represents a pair of links, one in either direction. The triangular nodes associate objects, properties and property values. Each node is an active element and when in an "active" state, sends out activation to all the nodes connected to it. A node may become active on receiving activation form another node in the memory network or a routine node. Triangular nodes behave slightly differently in that they become active only on receiving simultaneous activation from a pair of nodes.

The crude deroutineion given above is sufficient to demonstrate how simple recognition and retrieval tasks may be handled by such networks. To find the taste of HAM a routine would activate the nodes *HAS-TASTE* and HAM. The triangular node b1 linking *HAS-TASTE* and HAM to SALTY will receive coincident activation along two of its links and become active. As a result, it will transmit activation to SALTY which will ultimately become active. Figure 2 shows the activation levels of various nodes during the processing of the above query. If a routine needs to find an object that has a salty taste it would activate the nodes *HAS-TASTE* and SALTY. This will cause the same triangular node to become active and transmit activation to HAM. Eventually, HAM will become active completing the retrieval. The two examples roughly correspond to how retrieval and recognition may be processed by the network. In the rest of the paper we will focus on representational issues and hope that the example discussed above will give the reader an idea of the dynamics of network operation.

Representational framework

The semantic information forms a *conceptual structure* defined over a space spanned by *conceptual attributes.* All domain knowledge is defined in terms of these attributes and their values. Examples of attributes are: has-shape (with values such as round, triangular), has-color, is-an-instance-of and is-a-part-of.

The primary level of organization in the conceptual structure is in terms of *Concepts.* These are *labelled* clusters of "coherent" <attribute, value> pairs. The value of an attribute is also a Concept and hence Concepts may be arbitrarily complex. Concepts may refer to different sorts of things in the domain such as individuals, categories, events, properties, locations and relations. Attributes are classified into two broad categories: *PROPERTIES* and *structural links.* Properties correspond to the intrinsic features of Concepts and may vary from domain to domain. Thus, physical objects may have properties like *HAS-SHAPE* and *HAS-COLOR*, while events may have properties like *HAS-LOCATION* and *HAS-DURATION.* Structural links are fairly domain independent and define "inheritance-like" inference paths. The most representative of these is the *is-an-instance-of* link that is used to organize information in hierarchical structures in semantic networks. Our formulation employs an extended notion of property inheritance and includes other structural links such as the *is-a-part-of* and the *occurs-during* links [Allen 83] besides the *is-an-instance-of* link.

Concepts are classified into Types and Tokens. Tokens refer to instances and Types refer to abstractions defined over Tokens. Abstractions may in turn be defined over Types to yield more abstract Types, or a Type may be differentiated to result in more refined Types. In this framework, a Type is not viewed as a set and its structure is similar to that of a Token viz. a labelled collection of <attribute, value> pairs. The *is-an-instance-of* structural links encode the relation between a Token and a Type while *is-instantiated-by* links encode the inverse relationship.

We use a graphical notation for the representational framework. Figure 3 displays a sample network encoding the following information:

"Birds are a kind of Things, Swan is a kind of Bird, Hansa is a Swan, Things have the property color, Swans are generally White and White is a Color."

The representation uses three kinds of nodes: the Type node, the Token node and the Binder node. Arcs in the network represent bidirectional links. Type and Token nodes label clusters of <attribute, value> pairs, each of which is represented by a Binder node. For instance, b1 represents the fact: "Things have the property color", while b2 represents the fact: "Swans are generally colored white" i.e. "the value of the property color for Swans is generally White". The framework permits associating properties as well as property values with concepts. For example, we may represent that fruits have color without specifying any particular color values.

A weight is associated with each link and these provide the basis for the evidential semantics of knowledge. A link from node A to node B may be interpreted to mean "A provides evidence for B". Consider the links from Type nodes to their Binder nodes. The weights on these links provide a way of encoding the strength of generalizations represented by a Type. Thus, the link from SWAN to b1 in figure 3 is a quantitative measure of the evidence provided by the assertion "x is a Swan" to the assertion "the color of x is White". Cases with more than one typical value are easily represented as shown in figure 4. If red is a more typical color of Apple than green, the weight w1 will be greater than w2. The use of weights has other interesting consequences. For instance, if the node Apple is activated (the network is "imagining an Apple") activation from the node Apple will drive the associated Binder nodes. The Binder nodes corresponding to the most typical property values will receive the highest activation resulting in the activation of what would amount to a virtual Token corresponding to the most typical instance of the Type. Thus, the color of the imagined Apple will more likely be red than green. In this framework, the representation of a Type does double duty and acts as if it were a prototypical representation [Rosch 75], besides being an abstract representation of a class of Tokens.

The use of weighted links from a Type to its Binders provides a more natural interpretation of "exceptions" and "cancellations" and gives a clean semantics of the *is-an-instance-of* link. In this framework, one cannot both say: "All Swans are White" and "Giselle is a Swan whose color is black". However, one may say: "Most Swans are White" and "Giselle is a Swan whose color is black". This is illustrated in Figure 5. The crucial point is that Giselle may not be attached as an instance of Swan unless the weight of the link from Swan to b2 is reduced to a value less than 1.0. In Figure 5 the link from Swan to b2 is a statement of typicality and hence has a weight less than 1.0, whereas the link from Giselle to b3 encodes a definite statement and hence has a weight of 1.0.

Just as weights on links from concepts to Binders were significant, the weights on links from Binders to Concepts also serve an important function in categorizing an instance (assigning a Type to a collection of <attribute, value> pairs). The weights on links from Binders to Concepts can be used to assign a metric to the significance of a match between the <attribute, value> of a Type and that of an instance. The process of categorization easily translates into a "best fit" situation. Each Type receives evidence from Binders that match the input data. Type nodes accumulate this evidence and their level of activation provides a quantitative measure of the goodness of match. The Type with the highest activation wins [Feldman 82]. Furthermore, this also provides an interpretation of the notion of a prototypical instance of a category. If the property values of an instance match the typical values of the Type then the occurrence of this instance results in the higher activation of the Type node. Consequently, such an instance appears to be more prototypical. Thus, a Robin matches the properties in the representation of the Type Bird more strongly than a Penguin.

Conclusion

The representation and use of conceptual knowledge remains a core issue in cognitive science. This paper presents an approach to these problems that appears to offer several advantages over previous work. The basic ideas of evidential reasoning, multiple hierarchies and connectionist implementation fit together remarkably well and could form the basis for a detailed modeling of how knowledge is handled in natural systems.

References

[Allen 83] Allen James F. Maintaining knowledge about Temporal Intervals. *Commun.* ACM 26, 1983, 832-843.

[Fahlman 79] Fahlman Scott E. *NETL: A System for Representing and Using Real-World Knowledge.* The MIT Press, 1979.

[Feldman & Ballard 82] Feldman Jerome A., Ballard Dana H. Connectionist Models and their Properties. *Cognitive Science*, 6, pp 205-254, 1982.

[Feldman 82] Feldman Jerome A. Four Frames Suffice: A provisionary model of vision and space. TR99. Computer Science Department, University of Rochester, September 1982.

[Rosch 75] Rosch E. Cognitive Representations of Semantic Categories. In *Journal of Experimental Psychology*: General 104, 192-233, 1975.

[Shastri & Feldman 84] Lokendra Shastri, Feldman Jerome A. Semantic Networks and Neural Nets. TR 131. Computer Science Department, University of Rochester, January 1984.