ABOUT MODELING SOME ASPECTS OF HUMAN MEMORY

Boicho KOKINOv
Institute of Mathematics
Bulgarian Academy of Sciences
B1.8, Acad. G. Bonchev Street
Sofia 1113
BULGARIA

A computational model of knowledge representation in human memory is described. An attempt is made to combine frame representation with spreading activation. Mechanisms of inheritance, coreference, and retrieval are described. The model accounts for both forced and spontaneous retrieval. Some predictions are made and experimentally tested.

1. INTRODUCTION

An attempt to work out a computational model of knowledge representation in human memory is described. The model has been developed under the influence of many powerful ideas of Al researchers and psychologists, especially Minsky with his frame theory, Anderson with his spreading activation theory and Brachman with his structure inheritance nets. This work is a modification and further development of a model set forth by Barnev and KokinoV (3), (10).

The model has to account for some very different aspects of human cognition such as recalling and forgetting, perception of ambiguous stimuli (both as visual images and natural language phrases), reasoning, learning, etc. In developing the model all of these aspects have been taken into account but in the present paper only the model itself will be reported (Section 2) and the way it accounts for the retrieval aspects of human memory (Section 3). Two ways of testing the model are described as well (Section 4).

2. KNOWLEDGE REPRESENTATION IN THE MODEL

2.1. Overview

The representation of knowledge in the model may be viewed as a directed multigraph, i.e. a directed graph (network) with zero, one or more arcs between each two nodes; the nodes and the arcs (links) are labelled. The nodes correspond to descriptions of objects, concepts, events, images, programs for action, etc. We consider at least three types of
descriptions - modal specific (such as visual, auditory, tactile, etc. images), descriptions of motor actions, and conceptual descriptions (such as object, concept, and event descriptions, descriptions of mental actions, etc.). For the third type of descriptions, we propose a frame-like representation scheme; for now we put aside the question of representing modal information and motor programs, that we take into consideration the presence of such nodes in the network and we refer to them if necessary but we shall not consider their internal structure which may be frame-like as well or have some kind of analogical representation.

It should be stressed that words and concepts have separate descriptions in our model. This is unusual for most representation schemes; there are few exceptions and NETL (7) is one of them. Going further, we propose that words are completely separated from all kinds of conceptual description elements, i.e. the names of frames, slots, facets etc. are internal symbols for the system's own use which are invisible outside it. They are of course connected to the corresponding natural language words as well as to some kind of images or motor programs. The link which connects a concept or one of its elements to a word, image, etc. is called a m-coref link.

2.2. The Structure of a Frame

A frame (13) is a collection of declarative and procedural knowledge describing a conceptual unit such as a concept, object, relation, action, event, etc. and can think of a frame as a complex node with many links that enter or come out of it to connect it to other nodes. In fact knowledge is represented by these links, the node being merely a focus which can be referred to and where links cross. The links are represented as slots and facets in the frame. There are some slots common to all frames. The slot "type" with possible fillers such as "concept", "visual image", "word", "motor program", etc. indicates the type of description. The slots "is-a" and "instance-of" fix the type of the concept and carry generalisation and classification information. If there is an is-a slot, this means that this node describes an abstract concept such as a class of objects, a relation, some action, etc. and that this concept is a specialization of the concept pointed to by the filler. If there is an instance-of slot, this means that this node describes an individual concept such as an object, event etc. and that this concept is an instantiation of the concept pointed to by the filler. The slot m-coref points to the corresponding images, words, etc. whereas c-coref points to other conceptual descriptions of the same concept, i.e. the two descriptions link co-referent concepts and c-coref is a slot that describes an entity in the world. This allows for multiple descriptions of one and the same concept (13). The a-link will be discussed later. Besides these common slots each frame may contain an unlimited number of slots of its own. They may be classified into aspects, relations and actions. The aspects of a concept are its structural or functional parts (e.g. the hand, the foot in the notion of a man), abstract aspects (e.g. the occupation, the address in the notion of a man), or roles (e.g. the owner and the owned in the relation "possession"; the agent, the object, the instrument in the action concept, etc.). Relations are relationships between the aspects of the concept or between the concept itself and other concepts or their aspects. Actions are pieces of behavioral knowledge attached to this concept (e.g. how to recognize such a concept, how to act in the described situation, etc.). Every slot may contain different kinds of information that are represented by different facets.

The type facet defines the status of the slot: aspect, relation, or action. The is-a and instance-of facets define the slot as a specialization or respectively an instantiation of a slot of some generic concept. The c-coref facet points to a description (frame or slot) which refers to the same entity in the world. The m-coref facet points to an image or word description corresponding to this slot. The cardinality facet defines how many entities correspond to this slot. The modality facet gives some information to the interpreter (e.g. is the existence of an aspect obligatory or optional, is the holding of a relation a precondition for performing some action, etc.). The procedure facet points to some demons that are activated if some action is performed with the description. The a-link facet will be discussed later. The general structure of a frame is shown in Fig. 1.

![FIGURE 1](image)

The General Structure of a Frame

2.3. M-Coreference

The fact that natural language words are not used in frames makes the representation unambiguous, unique and language-free, and co-referent links connect frames and slots to corresponding natural language words. This allows for one and the same word to be linked to different frames and slots (homonymy) and vice versa one and the same frame or slot to be linked to different words (synonymy), including words from different languages.
2.4. C-Coreference

The c-coref link connects two conceptual descriptions (frames or slots) together expressing the fact that they both refer to the same entity in the world. This link helps us to solve many problems.

A concept may have different descriptions from different points of view, each of them represented in a separate frame (13). This increases the expressive power of the model (3) but raises problems such as how to determine whether two descriptions are of one and the same entity. This problem is unsolvable in syntactic and semantic terms. Only by using pragmatics (i.e., context, environment and goals) it is possible to find out that two descriptions refer to the same entity and then build a c-coref link between them. Now, using this link, it is no longer necessary to refer to the context and the environment. When connected to a slot the c-coref link plays the role of a pointer either to the value or to the range of the slot. Whether it will be the former or the latter depends on the instance-of and the is-a facets of the same slot. The internal names of the slots are local, i.e., we can refer to them only by referring to the corresponding frame. Frames that are specializations or instantiations of another frame need not have the same slot names as their ancestors. For that reason when, for example, Slot1 of frame G1 corresponds to Slot5 of frame G6 (see Fig. 2), then they must be connected by an is-a or instance-of link. If Slot1 is a specialization of Slot5 (there is an is-a link between them), then c-coref facet of Slot1 points to its range which may be a subrange of the range of Slot5. If Slot1 is an instance of Slot5 (there is an instance-of link between them), then c-coref points to the value of Slot1 which must belong to the range of Slot5. In this way slot values and restrictions can be modified.

G1
is-a: (G6 1) slot1
is-a:((G6 slot5) 1) c-coref:((G5 1))

G1
is-a:((G6 slot5) 1) slot1
is-a:((G6 slot5) 1) c-coref:((G5 1))

FIGURE 2
C-coref link replaces the value and range links

Furthermore, slot values can be differentiated when two or more slots in the successor are connected to one and the same slot of the ancestor (the generic frame) whose cardinality allows this (see Fig. 3 for an example).

<table>
<thead>
<tr>
<th>Slot1</th>
<th>Slot2</th>
</tr>
</thead>
<tbody>
<tr>
<td>type: aspect</td>
<td>type: aspect</td>
</tr>
<tr>
<td>is-a: (G7 slot5) 1</td>
<td>is-a: (G6 slot5) 1</td>
</tr>
<tr>
<td>c-coref: (clothes 1)</td>
<td>c-coref: (shirt 1)</td>
</tr>
<tr>
<td>m-coref: (shirt 0.6)</td>
<td>m-coref: (trousers 0.6)</td>
</tr>
<tr>
<td>cardinality: (0 20)</td>
<td>cardinality: (1 1)</td>
</tr>
</tbody>
</table>

FIGURE 3
Differentiation of an ancestor slot

And finally this representation allows implementing one of the most exciting ideas of Minsky's famous paper (13) - the idea of frame systems (or frame arrays as called in his new book (14)). A frame array is a group of frames that share the same terminals. This makes the recalculation of the same values for the same terminals unnecessary when the viewpoint changes from one of these frames to another. As Minsky (14) stated, this idea did not become popular. One of the reasons for that is, according to us, that in most frame-based systems slots are local and therefore they could not be shared. In our representation scheme only the internal name of the slot is local but its value or range are not (as pointed out before). They are connected by a c-coref link to the slot. That is why using our representation scheme it is possible to implement a frame array (see Fig. 4).

2.5. A-Links

All the links considered so far have a strong equistemological interpretation - specialization, individualization, part-of, etc. Very often, however, we form associative links in our memory between arbitrary (from an epistemological point of view) things, e.g., between two events that have occurred within a short period of time, or between characteristic features that are often found together, etc. We call this kind of link an a-link; such links are used only to facilitate the access to the stored information, i.e., to form our preliminary setting. A-links are formed between frames as well as between slots, on one hand, and frames or other slots, on the other hand. The role of a-links will be discussed in more detail in the next section.

3. RETRIEVAL

The role of the retrieval process is crucial for all
G1
instance-of: (scene)
c-coref: (G2 1)
slot1
  type: aspect
  instance-of: (G6 1)
m-coref: (chair 1)
c-coref: (G10 1)
slot2
type: aspect
  instance-of: (G9 1)
m-coref: (table 1)
c-coref: (G11 1)

G2
instance-of: (scene)
c-coref: (G1 1)
slot1
  type: aspect
  instance-of: (G6 1)
m-coref: (chair 1)
c-coref: (G10 1)
slot2
type: aspect
  instance-of: (G9 1)
m-coref: (table 1)
c-coref: (G11 1)
slot3
type: relation
  instance-of: (G8 1)
c-coref: (G21 1)

G21
instance-of: (G25 1)
m-coref: (near-by 1)
slot1
type: aspect
  instance-of: (G20 slot1 1)
c-coref: (G1 slot1 1)
slot2
type: aspect
  instance-of: (G20 slot2 1)
c-coref: (G1 slot2 1)

Figure 4: A frame array describing the transformation of a scene.

A mental activities because the knowledge base stored in human memory is very large. The exhaustive search of memory is impossible. That is why many cognitive science researchers have paid great attention to building efficient mechanisms for knowledge retrieval. Schank (16, 17, 18), Kolodner (11, 12), and Carbonell (6) have considered the organization of memory very significant for all reasoning processes and have built up models of that organization. They have proposed a way of organizing knowledge called MOPS or EMOPS and a variety of indexing mechanisms for efficient retrieval of the required knowledge (events, generalized events, scripts, etc.) which is of crucial significance for the case-based reasoning developed by the mentioned authors. A shortcoming of this approach is that once indexed in a definite way, knowledge can be retrieved only by means of this index, i.e., in a static way.

Anderson (1, 2) has developed the old idea of spreading activation and has build a model of human memory based on a version of the latter. In his model, declarative knowledge is represented by a semantic network whereas procedural knowledge - by a production system. The basic idea is that nodes in the network vary in terms of their states, i.e., they have various levels of activation, and the speed with which they will be processed by the production system depends on their activity. The more active the node is, the more early will be the attempt to process it. Activation spreads automatically from a node to each of its neighbors depending on the strength of the link between them (calculated dynamically). Holland, Holyoak et al. (8) have proposed a directed version of spreading activation. Their model is totally rule-based and there are rules (called associative rules) that cause the activation of a concept. In this was, they accounted for the strategic aspects of human retrieval.

We share Rau's opinion (15) that both forced and spontaneous retrieval exist in human memory. Some experimental results (19) support this idea; that is why we make an attempt to cover both aspects of retrieval in our model.

All frames have their activation levels and the activation can spread from one frame to another by means of all kinds of links between them (is-a, instance-of, c-coref, m-coref, a-link, etc.). All links have their own strengths (between 0 and 1) and the activity handed over the link is proportional to its strength. All nodes with a level of activity which exceeds a definite threshold are in the active working memory (WM) and the most active one is called focus of WM. All nodes in WM pass their activation over to their neighbors by means of the associative mechanism (3) which is a version of Anderson's spreading activation. This is done automatically at the unconscious level. The only node that can be processed consciously is the focus and if some information must be obtained from a node referred to by the focus then the activation has to be directed to that node via the corresponding link (is-a, instance-of, c-coref, or another). Usually these links have the strength of 0 and therefore the node pointed to by the link will receive the same activation as the focus and thus become the focus itself; so it will be possible that it is processed consciously.

This mechanism will be called directed spreading activation and also forced or strategic retrieval. In parallel with this process, however, all nodes in WM pass their activation over to their neighbors, they in their turn pass it out to their neighbors and so on (this is the automatic or spreading activation mechanism). If some node receives more activation during this process than the current focus, then obviously this node will become the focus; this phenomenon is called spontaneous retrieval. Usually the conscious mechanism dominates the unconscious, that is, strategic retrieval suppresses the result of spontaneous retrieval because of the greater level of activation in the focus. In some situations, however, spontaneous retrieval may dominate forced retrieval, for example:

- when the focus is not clearly expressed, i.e., when there are nodes whose activity is almost as great as that of the focus;
- when the link along which activation is directed is broken (i.e., it connects the focus to several branches) and different nodes each branch having its own strength, and activity is distributed among these links;
- when the attention is not focused upon a definite slot or relation;
- during sleeping and dreaming.
So this single mechanism of spreading activation called associative mechanism can account for both forced and spontaneous retrieval.

4. TESTING THE MODEL

Knowledge representation models of human memory can not be tested directly, but there are at least two indirect ways for doing this. Firstly, model predictions for some behavioral phenomena can be tested by a psychological experiment. The model under consideration makes some predictions related to the influence of the preliminary setting upon fact retrieval and problem solving. Psychological experiments are planned to explore this and one of them is reported in section 4.1. Secondly, computer simulations based on this knowledge representation scheme can be done. The behavior of the computer system can be compared to that of humans under the same conditions.

4.1. Psychological Experiment

According to our model, to recall some facts means to cause the node describing these facts to become the focus and to forget means to be unable to activate the corresponding node. If this is the case then if we activate nodes closely connected with the node to be retrieved before a question is asked then the retrieval time can be expected to be shorter. We have planned the following experiment in order to test this hypothesis.

4.1.1. Method

Materials: Questions are asked about well known but not very common and familiar facts on geography, chemistry, Bulgarian history. Facts must not be very familiar to the subjects because such facts are reported very quickly and the answers will not be very sensitive to external influence. (For this reason some of the subjects did not actually know the answers to the questions they were asked.) The preliminary setting was achieved by means of supplying the subjects with some preliminary information that was to activate nodes closely related to the one to be retrieved. The questions together with the preliminary information are shown in Table 1.

Subjects and Apparatus: The subjects were pupils and undergraduate students in mathematics. Only the subjects who turned out to know the correct answer to the questions were taken into consideration and that is why their number for each question and each condition may be different. The experiments were carried out using a self-made computer system running on an IBM PC/XT.

Procedure: Subjects were asked questions being instructed that before the questions they may receive some preliminary information either relevant or irrelevant to the question. That information and the questions were displayed to the subjects on the computer terminal and they reported the answers aloud, the experimenter having to register the reaction time by pressing a key on the keyboard. Each subject received questions with or without preliminary information in an arbitrary order. Each question was accompanied by preliminary information for about half of the subjects.

Results are shown in Table 2. The differences between the mean reaction time (RT) for the compared conditions are significant at the 5% level measured by Kolmogorov-Smirnov's criterion. The differences between the mean RT under different conditions for the third question are significant at the 1% level. We can also compare the mean RT for question 3 and question 4 without preliminary setting for both. The former is 7.575 sec and the latter is 4.768 sec, \( \chi^2 = 2.48, \ p (0.05) \). We think that when asking the same question by different words and phrases different nodes are activated from which starts the retrieval process.

<table>
<thead>
<tr>
<th>question</th>
<th>with preliminary setting</th>
<th>without preliminary setting</th>
<th>( \chi^2 )</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of subjects</td>
<td>mean RT</td>
<td>number of subjects</td>
<td>mean RT</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>3.159</td>
<td>12</td>
<td>5.622</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>3.987</td>
<td>8</td>
<td>8.049</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>4.594</td>
<td>12</td>
<td>7.575</td>
</tr>
</tbody>
</table>

4.2. Computer Simulation.

An older version of the model was implemented in IQ Lisp and experiments were made with it within an
information-logical system. Simulations of fact retrieval and deductive reasoning were performed. The dialogue with the system was in a restricted natural language using polysemic words. Simulations of a free association experiment were also performed. It was a small implementation which gave encouraging results (9). An implementation of the ideas developed so far in an attempt to model common-sense reasoning is now under development. It is written in Golden Common Lisp for the IBM PC/XT computer.

5. CONCLUSIONS

The model under consideration is an attempt to combine the frame idea proposed by Minsky (13) and the spreading activation idea proposed by Anderson (1). A frame representation scheme is proposed which develops further some of Brachman's ideas (4, 5) on epistemological links. Traditional facets such as value and range are replaced by the c-core links which supply a more powerful and flexible mechanism. Slot inheritance is developed to be more powerful by assigning to each slot is-a and instance-of facets as well. Frame representation is fully liberated from natural language closure and m-coref links are introduced to connect concepts to words and images allowing them to be perceived and expressed in the external world. A new kind of link, namely the a-link, is introduced to express some more arbitrary associations which facilitate the retrieval process.

A retrieval mechanism is proposed which makes an attempt to account for both aspects - forced and spontaneous - of memory search. A psychological experiment is described which tests the model's predictions in a fact retrieval paradigm.

ACKNOWLEDGEMENTS

I am grateful to Peter Barnev and to all participants in the interdisciplinary seminar on Cognitive Science at the Institute of Mathematics, B.A.S. for their discussions relevant to the model and especially to Stoyanka Zlateva and Encho Gerganov for their valuable remarks on a draft of the paper. I would like to thank Encho Gerganov for his help in planning the experiment, Vassil Nikolov and Ivan Bosov for implementing the computer system supporting it. I would like to express my special thanks to Vassil Nikolov for his great help in performing the experiments and in the text preparation. This work was partially supported by Contract No 607 with the Committee for Science.

REFERENCES

(10) Kokinov, B., Computational Model of Knowledge Organisation and Reasoning in Human Memory, in Mathematics and Education in Mathematics (BAS, Sofia, 1988).