

FUZZY SETS IN KNOWLEDGE ENGINEERING

C. V. Negoita

ABSTRACT

Any representation scheme for building a knowledge-based system is embedded in a topos whose logic determines the inference mechanism.

1. INTRODUCTION

The past two decades have seen the development in Artificial Intelligence research of a class of computer programs that can reach an expert level of performance. Termed knowledge-based systems, they have demonstrated ability to solve complex problems in a broad assortment of applications ranging from medicine to finance. As embedded systems they will serve as the control systems of many sophisticated machines.

Historically, knowledge engineering is a product of post-World War II ferment in cybernetics, decision sciences, and artificial intelligence. A question that has confronted the computing field since its inception is how machines can be made smarter. Any traditional computer program directs the machine to access data, but the decisions about how to process those data are invariably coded. These decisions are made by a programmer.

Knowledge engineering is a discipline devoted to integrating human knowledge in computer systems. The distinctive characteristic of any knowledge-based system is that its processes are state-driven rather than hard-coded. Decisions about how to process data are included in the system as pieces of knowledge that the system digests. In other words, an intelligent system is writing its program.

To know means to internalize, to have in mind as the result of experience or of being informed, or because one has learned. In what follows knowledge will be defined as internalized instructions on how to perform in certain situations, what helps the computer to do smart things.

The field of knowledge engineering is split into two camps. The "engineers" are trying to get their programs to do smart things. The scientists, a much smaller group, are after a general theory of intelligence, both human and machine. The fundamental assumption of Knowledge Engineering is that mind can be modeled as a processor of symbols - in effect, as a computer program. Cognition is considered to be a high-level process, which means that it can no more be understood in terms of the firing of individual neurons than a computer program can be understood in terms of the 1's and 0's filtering through an individual memory register (Waldrop 84).

The main problems of knowledge engineering are acquisition, representation, and inference. They are so interrelated that one considers them as three faces of the same problem. There seems to be something very profound in this interrelation and it is time for theoretical investigations.

Finally, there is the matter of reasoning with uncertainty. In the case of plausible reasoning, uncertainty is supposed to be in the state of our knowledge, and certainty factors are assigned to facts and rules in order to generate an estimate of reliability. Efforts have been made to apply probability theory, without much success. Most of the time, the uncertainty in our knowledge is less a question of randomness than of vagueness.

In approximate reasoning, uncertainty is supposed to be an intrinsic property of natural language (Zadeh, 83). It is the goal of this paper to show that the difference between plausible and approximate reasoning can be better understood via an algebraic framework.

## 2. THINKING IN RULES

Knowledge can be represented in a variety of ways including production rules and first-order logic. Production rules have usually the form

IF condition THEN action

where the antecedent represents some pattern, e.g. a conjunction of predicates, and the consequent specifies an action to be taken when the data match the pattern. For instance:

- (1) IF stain is positive THEN organism is streptococcus
- (2) IF organism is strep THEN penicillin is indicated

A production system is a list of rules

IF A THEN B

IF B THEN C

IF C THEN D

which can be represented as arrows

$A \longrightarrow B$

$B \longrightarrow C$

$C \longrightarrow D$

Each arrow is associated with two objects and an operation of composition can be performed:

$A \longrightarrow B \longrightarrow C \longrightarrow D$

Clearly, any production system is embedded (fixed firmly) in a category (Negoita 84), a universe of a particular kind of representation. Such a universe is determined by specifying a certain kind of object, and a certain kind of arrow that links the objects. Thus, the study of any production system can be developed according to the category modeling the structure of the knowledge base.

Much representation of the world is concerned with conceptual relations expressed in sentences such as:

stain is positive  
organism is streptococcus

Two predominant forms of statements are predications, expressing that an individual (e.g. stain) is of certain type (e.g. positive) and the universally quantified conditional, expressing that one type (e.g. organism) is a subtype of another (e.g. strep).

One of the primary perspective offered by the theory of categories is that the concept of arrow may be used instead of the membership relation as the basic building bloc expressing properties. Instead of defining properties of a collection by reference to its members, i.e. internal structure, one can proceed by reference to its external relationships with other collections. The links between collections are provided by arrows, and the axioms for a

category derive from the properties of arrows under composition.

### 3. EVIDENCE COMBINATION

A number of standard diagrams can be formulated in the language of arrows to model the mechanisms of evidence combination. The central idea is that of a cone of a diagram, a new object induced by a diagram.

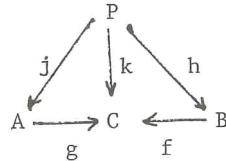
Consider two production rules:

$$\begin{array}{l} A \longrightarrow C \\ B \longrightarrow C \end{array}$$

and the diagram

$$A \longrightarrow C \longleftarrow B$$

A cone for this diagram consists of a new object  $P$  and three arrows,  $h$ ,  $j$ ,  $k$ , such that



commutes. This is a pullback square, and we say that  $h$  arises by pulling back  $f$  along  $g$ .

Clearly, evidence combination can be represented by a diagram in the category of facts and rules, and the meaning of the new object  $P$  is  $(A \text{ and } B)$ . We say that the operator AND is internalized in the pullback square, or described by it. The nature of this operator is determined by the nature of the category where facts and rules are represented as objects and arrows. We will see further that the nature of the category determines the nature of logic underlying the pullback construction and any other construction based on the idea of a cone.

### 4. SETS AND GENERALIZED SETS

The recent, though long predicted, attention and the ongoing detailed study of fuzzy sets in a special category called topos (Stout 84), open the door for a qualitatively new type of knowledge diagrams and may provide a major new tool for knowledge engineering. For instance, if predication

John is tall

is represented as a function

$$\text{tall: heights} \longrightarrow [0;1]$$

then we can investigate knowledge diagrams in the category of fuzzy sets. This function is a table

$H_1$	$h_1$
$H_2$	$h_2$
$H_3$	$h_3$
$H_4$	$h_4$
...	...

and according to the theorem of representation (Negoita, Ralescu 75) such a table is equivalent with another one

$H_1$
$H_1 H_2$
$H_1 H_2 H_3$
$H_1 H_2 H_3 H_4$
.....

which can be seen as a generalized set having partial elements characterized by degrees of existence, or sets-through-time

$$H_1 \longrightarrow H_1 H_2 \longrightarrow H_1 H_2 H_3 \longrightarrow H_1 H_2 H_3 H_4 \longrightarrow \dots$$

A generalized set is a set-like object, and their category has a structure in which the interpretations of basic set-theoretical constructions behave much as they do in the category of sets (Negoita, Stefanescu 82).

The category of generalized sets is a topos. Topoi are of interest to logicians because they have an internal logic which allows higher order theories to be modeled in them. Since almost all of knowledge engineering can be phrased in higher order theories this means knowledge engineering can be modeled in a topos. Fuzzy mathematics uses fuzzy sets for the same kind of modeling, and embedding fuzzy sets in topoi, things become more clear.

## 5. KNOWLEDGE DIAGRAMS

Modeling a knowledge base as a collection of objects and arrows, leads to a new interpretation of the inference mechanisms. As an example, consider a knowledge base

```
LIKES(Name1 → Name2)
PERSON(Name2 → Heights)
TALL(Heights → 0.1)
```

and a query

```
WHO LIKES TALL PERSONS?
```

The answer by composing three relations according to the pullback diagram

$$\begin{array}{ccc} \text{Answer} & \longrightarrow & \text{TALL} \\ \downarrow & & \downarrow \\ \text{LIKES} & \longrightarrow & \text{PERSON} \end{array}$$

The same type of diagram models the syllogism:

```
IF A THEN B
A*
```

which can be represented as

$$\begin{array}{ccc} A^* & \longrightarrow & \text{Answer} \\ \downarrow & & \downarrow \\ A & \longrightarrow & B \end{array}$$

## 6. CONCLUSION

The posing of a question to a knowledge base is a forward process resulting in the generation of a knowledge tree. The answer of a question to a knowledge base is a backward contraction process in the tree, and this process can be modeled by a pullback in the category of representations.

## REFERENCES

- [1] Negoita, C. V. *Fuzzy Systems and Expert Systems*. Menlo Park, CA: Benjamin/Cummings, 1984).
- [2] Negoita, C. V. and Ralescu, D. A. *Applications of Fuzzy Sets to System Analysis* (New York: Halsted Press, 1975).

- [3] Negoita, C. V. and Stefanescu, C. *On Fuzzy Optimization: Fuzzy Information and Decision Processes*, M. Gupta and E. Sanchez, eds. (Amsterdam: North-Holland, 1982).
- [4] Stout, L. N. "Topoi and categories of fuzzy sets", *Fuzzy Sets and Systems* 12 (1984), 169-184.
- [5] Waldrop, M. M. "The necessity of knowledge", *Science* 223 (March 1984), 1279-1282.
- [6] Zadeh, L. A. "The role of fuzzy logic in the management of uncertainty in expert systems", *Fuzzy Sets and Systems* 11 (1983), 199-227.

