Fuzzy reasoning in a multi-dimensional space of hypotheses

R. Kruse
Institut für Betriebssysteme und Rechnerverbund
Technische Universität Braunschweig

E. Schwecke
Institut für mathematische Stochastik
Technische Universität Braunschweig

Abstract: Aim of this paper is to provide a basic mathematical model for fuzzy-reasoning systems. Within this framework, the generalized modus ponens and modus tollens are formalized, and a mathematical concept of evidence is introduced. It is assumed, that numerical representations of all fuzzy-informations are available – the problem of linguistic modelling is not investigated.

Keywords: fuzzy-reasoning, approximate reasoning, uncertainty, evidence

1. Introduction

In the last decade the treatment of uncertainty has been perceived as one of the important problems in the field of knowledge based systems. Based on the fuzzy-set theory L.A. Zadeh developed in 1977 the theory of approximate-reasoning which allows the generalization of modus ponens and modus tollens – two basic concepts of the classical logic. Drawing inferences by the use of the generalized modus ponens means to conclude $Y$is C' by the application of a rule like IF X is A THEN Y is C to an antecedent X is A.

In [ZADE 77] as well as in [MIZI 84] and many other publications the rule-representation and the inference-procedure are considered as two-dimensional problems and the propagation as a sequence of seperate rule-applications where every conclusion is antecedent for the next step.

In the following chapters a mathematical model for the entire reasoning process is introduced. Within this model the considered 'part of the world' is represented by a number of characteristics (e.g. age, weight, height of a person) and their values (e.g. 27 years, 92 kg, 182 cm). The set Ω contains all 'possible states of the world' and a relation R on Ω describes the causal dependencies of the characteristics. Since Ω has a multi-dimensional structure, concepts like modus ponens and modus tollens can be interpreted as two-dimesional projections of multi-dimensional problems. From this point of view the sequential use of rules has to be considered as a sequential projection of one multi-dimensional causal inference to different projection planes. Starting point of every propagation process are user observations. Their integration in the mathematical model requires a formal concept of evidence, embedded into the theoretical framework mentioned above.

Based on the results of the mathematical modelling a propagation algorithm can be presented, which allows to perform the propagation by local computations on communicating parallel working node-processors. In the range of probabilistic reasoning the idea of a multi-dimensional space of hypotheses and propagation by message-exchanging node-processors was inspired by the work of [PEAR 86]. Our intention is to formulate an analogous mathematical model for fuzzy-reasoning which allows a formal definition of concepts like evidence and propagation.

2. Reasoning with ordinary sets

To clarify the basic ideas and intentions we will restrict ourselves to ordinary sets, but since all notations and concepts introduced in the following are chosen under the aspect of a later generalization the extension to fuzzy sets causes no difficulties. At first we have to consider the representation of subsets A of a finite discrete set Ω by its indicator-function $I_A : \Omega \rightarrow \{0,1\}$. The set-operations like union, intersection, complement, projection and cylindrical extension can be formulated as operations with indicator-functions.

2.1. The basic mathematical model

The characteristics of interest may be denoted as $X^{(1)},\ldots,X^{(m)}$, the sets

$$\Omega^{(i)} = \{\, x_1^{(i)},\ldots,x_{n_i}^{(i)} \,\}, \quad i \in \{\, 1,\ldots,m \,\}, \tag{1}$$

contain their possible values. The set of 'all possible states of the world' is defined by the cartesian product

$$\Omega \stackrel{\mathrm{d}}{=} \mathop{\times}_{i=1}^{m} \Omega^{(i)}. \tag{2}$$

(In the sequel Ω^N with $N \subseteq \{\, 1,\ldots,m \,\}$ denotes the cartesian product of the sets $\Omega^{(k)}$, $k \in N$. As a special case we get $\Omega = \Omega^{\{1,\ldots,m\}}$.) The characteristics $X^{(i)}$ are interpreted as set-variables; their values are subsets of $\Omega^{(i)}$ which restrict the actual but unknown values of the characteristics.

A relation $R \subset \Omega$ summarizes the expert-knowledge of all causal dependencies. The elements of R are defined by non-contradictory combinations of properties, the projection of R to $\Omega^{(i)}$ leads to an a-priori-restriction for the characteristic $X^{(i)}$.

If for example the dependencies of $X^{(1)}$ (age) and $X^{(2)}$ (*weight*) with $\Omega^{(1)} = \{\ 1,\ldots,100\ \}$ and $\Omega^{(2)} = \{\ 3,\ldots,120\ \}$ are considered, elements of Ω like (27, 15) will surely not be elements of the relation R, since every person of age 27 will weigh more than 15kg.

2.2. The concept of evidence

Aim of any reasoning process is to draw inferences to unobserved characteristics from the obsereved values of other accessible ones. Observing in our model means the ability to restrict the set of possible values $\Omega^{(i)}$ to $D^{(i)} \subset \Omega$. In the best case this restriction $D^{(i)}$ contains only the one actual value of the i-th characteristic, in the worst case we have $D^{(i)} = \Omega^{(i)}$ and nothing can be said about the actual value of the i-th characteristic. The cylindrical extensions of the observations $D^{(i)}$ are denoted as

$$E^{(i)} \stackrel{\mathrm{d}}{=} \operatorname*{cyl}_{\Omega} D^{(i)} . \tag{3}$$

This subsets of Ω are called the 'evidences' caused by the observations '$X^{(i)}$ is $D^{(i)}$'. Each set $E^{(i)}$ contains those combinations of properties, which are consistent with the observation of $X^{(i)}$. The collection of several observations is done by the intersection of their cylindrical extensions. If we define $D^{(j)} \stackrel{d}{=} \Omega^{(j)}$ for unobserved characteristics we derive

$$E \stackrel{d}{=} \bigcap_{i=1}^{m} E^{(i)} \ , \tag{4}$$

where E is the total evidence. As R summarizes the expert knowledge E summarizes the user-observations; only $R \cap E$ is consistent with both, the expert knowledge and the user-observations. The projection of $R \cap E$ on $\Omega^{(k)}$, where $X^{(k)}$ is the characteristic of interest, yields the restriction of $X^{(k)}$ which can be concluded from the current observations.

2.3. How to get the relation

Since in reality all characteristics under consideration are in interaction whith each other, the causal dependencies of two selected properties depend on the values of the remaining attributes. Neglecting this mutual inferences the two-dimensional relations $R^{(i,j)}$ on $\Omega^{\{i,j\}}$ can be obtained by the projection of R to the projection-plane $\Omega^{\{i,j\}}$.

$$R^{(i,j)} \stackrel{d}{=} \operatorname*{proj}_{\Omega^{\{i,j\}}} R . \tag{5}$$

Although this is an idealizing assumption, by the use of the projections $R^{(i,j)}$ no consistent elements of Ω are lost, but the consequence is a descending strength of the restrictions. In the worst case there are only trivial results like $X^{(i)}$ is $\Omega^{(i)}$.

The problems concerning the application of the model are conversely – the relations $R^{(i,j)}$ describing the dependencies of special pairs of characteristics are known and the m-dimensional relation R has to be constructed. Assuming independence in the sense mentioned above this is done by intersecting the cylindrical extensions. For those relations which are not specified by the expert we define $R^{(i,j)} \stackrel{d}{=} \Omega^{\{i,j\}}$. We get

$$R = \bigcap_{\{i,j\} \subset \{1,\dots,m\}} \left(\operatorname*{cyl}_{\Omega} R^{(i,j)} \right) ; \tag{6}$$

this relation will be a superset of the 'true' relation R^*.

The expert has to provide two-dimensional relations $R^{(i,j)}$ for a subset of the $\binom{m}{2}$ possible projection-planes. Those projection-planes in which the relations $R^{(i,j)}$ are specified by rules like

$$\text{IF } X^{(i)} \text{ is } A^{(i)} \text{ THEN } X^{(j)} \text{ is } A^{(j)}, \quad A^{(i)} \subseteq \Omega^{(i)},\ A^{(j)} \subseteq \Omega^{(j)},$$

we call rule-planes.

Rules of this kind are generalizing what we know as 'modus ponens'. Such rule expresses that from the fact $X^{(i)}$ is $A^{(i)}$ it can be concluded $X^{(j)}$ is $A^{(j)}$; but if the value of $X^{(i)}$ lies in $\overline{A^{(i)}}$ nothing can be concluded. Nevertheless every rule is bidirectional in the sense of modus tollens, knowing $X^{(j)}$ is $\overline{A^{(j)}}$ it can be inferred $X^{(i)}$ is $\overline{A^{(i)}}$ [ZADE 77].

A relational representation of IF ... THEN ...-rules can be consructed by the following expression:

$$R^{(i,j)} = \{ \underset{\Omega^{\{i,j\}}}{\text{cyl}} (A^{(i)}) \cap \underset{\Omega^{\{i,j\}}}{\text{cyl}} (A^{(j)}) \} \cup \underset{\Omega^{\{i,j\}}}{\text{cyl}} \overline{(A^{(i)})}. \tag{7}$$

Unfortunately in most cases one single rule will not be sufficient to represent the total knowledge about the dependencies of two characteristics $X^{(i)}$ and $X^{(j)}$. The expert will formulate several rules in 'both directions' like

IF $X^{(i)}$ is $A_1^{(i)}$ THEN $X^{(j)}$ is $A_1^{(j)}$

IF $X^{(j)}$ is $A_2^{(j)}$ THEN $X^{(i)}$ is $A_2^{(i)}$

$\vdots$

Every rule yields a relation on $\Omega^{\{i,j\}}$ which contains consistent pairs of values. To obtain the overall relation $R^{(i,j)}$ we have to intersect the different partial relations. In this way all rules concerning the characteristics $X^{(i)}$ and $X^{(j)}$ can be combined.

2.4. Problems of realization

Considering 10 characteristics with only 5 possible values for each requires the handling of an 10-dimensional space Ω with 9,765,625 possible 'states of the world'. This shows, that even relatively trivial constellations lead to impracticable problems concerning the handling of R. The high-dimensional propagation-process has to be splitted into several two-dimensional problems. Instead of constructing firstly R by the two-dimensional relations $R^{(i,j)}$ and second E by the observations $D^{(i)}$ to intersect both at last, we carry out the intersection at the rule-plane level, so the algorithm yields the the projection of $R \cap E$ to the axis of interest without the necessity of handling R or E as a whole.
The strategy is to represent the characteristics $X^{(1)}, \ldots, X^{(m)}$ and their causal dependencies by a network of node- and link-processors, exchanging impulses and messages. At each node $X(i)$ current restrictions $B^{(i)}$ are stored and modified during the propagation process. When the algorithm stops this restrictions $B^{(i)}$ equal the projections of $E \cap R$ to the sets $\Omega^{(i)}$.

2.5. The algorithm

Each characteristic $X^{(i)}$ is represented by a node-processor which stores the current restriction $B^{(i)}$ of $\Omega^{(i)}$. For every rule-plane we have a link-processor $R^{(i,j)}$, providing the relation with the same name. $R^{(i,j)}$ is connected with $X^{(i)}$ and $X^{(j)}$. The current-restriction-sets $B^{(i)}$ are initialized with $\Omega^{(i)}$ and modified, if the node-processor receives messages $B^{(i|j)} \subseteq \Omega^{(i)}$ from neighbouring link-processors or external messages $B^{(i|0)}$ representing observations $D^{(i)}$. If these messages are not subsets of $B^{(i)}$ the new value of $B^{(i)}$ is defined by $B^{(i)} \cap B^{(i|j)}$, in the other case no updating takes place. After every modification of a restriction-set $B^{(i)}$ the node-processor sends impulses to all neighbouring link-processors $R^{(i,j)}$.

These impulses cause the node-processor to calculate the messages

$$B^{(i|k)} = \operatorname*{proj}_{\Omega^{(i)}} \left(\operatorname*{cyl}_{\Omega^{\{i,k\}}} B^{(i)} \cap \operatorname*{cyl}_{\Omega^{\{i,k\}}} B^{(k)} \cap R^{(i,k)} \right),$$

$$B^{(k|i)} = \operatorname*{proj}_{\Omega^{(k)}} \left(\operatorname*{cyl}_{\Omega^{\{i,k\}}} B^{(k)} \cap \operatorname*{cyl}_{\Omega^{\{i,k\}}} B^{(i)} \cap R^{(i,k)} \right).$$

and to send them to the directly connected node-processors. If the flew of impulses and messages ends since all restrictions remain unmodified, the sets $B^{(i)}$ equal the projections of $E \cap R$ to the characteristics $X^{(i)}$.

3. Fuzzyfication

The extension of the mathematical model and the algorithm introduced in the previous chapters is essentially done by the extension of two-valued indicator-functions to real-valued membership-functions.

Problems arise from the construction of fuzzy relations. Several different solutions are suggested, their advantages and deficiencies are investigated e.g. in [MIZI 84]. The straightforeward extension of (7) leads to the procedure proposed in [ZADE 77].

For the implementation of the algorithm on a digital computer the representation of fuzzy sets by level-sets [KRME 87] is useful. With the exception of the complement all set-operations used in the algorithm can be lead back to corresponding operations with the (crisp) level-sets. This allows to split a fuzzy-propagation-process into a small number of independent crisp propagation-processes.

4. Concluding remarks

In this paper we introduced a mathematical model for fuzzy reasoning-systems. It is founded on a multi-dimensional space of possible 'states of the world' and allows to get formal definitions of concepts like evidence and propagation. The entire reasoning process can be validated since a formal mathematical framework is provided. One result of the exertions is a simple propagation-algorithm which can easily be implemented in an object oriented environment. It has to be emphazised, that several serious problems in the field of approximate reasoning like linguistic modelling are neglected.

5. References

[KRUS 86] Kruse, R., 'Uncertainty-Representation and Uncertainty-Propagation in Expert Systems', *Technical Report* , Braunschweig 1986.

[KRME 87] Kruse, R. and Meyer, K.D., 'Statistics with Vague Data', *D. Reidel*, Monograph, Dordrecht 1987.

[MIZI 84] Mitzumoto, M. and Zimmermann, H.J., 'Comparism of Fuzzy-Reasoning-Methods', *Fuzzy Sets and Systems* 8, 253–283, 1984.

[PEAR 86] Pearl, J., 'Fusion, Propagation, and Structuring in Belief-Networks', *AI* 29 (13), 241–288, 1986.

[ZADE 77] Zadeh, L.A., 'The Concept of a Linguistic Variable and its Application to Approximate Reasoning', *Information Sciences* 8, 198–249, 301–357 and *Information Sciences* 9, 43–80, 1977.

[ZADE 79] Zadeh, L.A., 'A Theory of Approximate Reasoning' in *Machine Intelligence* 9, Hayes, J.E., Michie, D. and Mikulich, L.I. (eds.), Wiley, 149–194, 1977.