Sborník vědeckých prací Vysoké školy báňské - Technické univerzity Ostrava
číslo 2, rok 2005, ročník LI, řada strojní
článek č. 1466

Štefánia GALLOVÁ*

UNCERTAINTY AND VAGUENESS OF REASONING

NEURČITOSŤ V ROZHODOVANÍ

Abstrakt

Príspevok je zameraný na jeden z metodologických postupov riešenia diagnostických problémov v prostredí znalostného (expertného) systému, ktoré umožňuje v dostatočnej miere aplikovať princípy umelej inteligencie tak, aby bolo možné pracovať aj s nepresnými, vágnymi, neurčitými, jako aj nejednoznačnými informáciami, ktoré reálna prax v prevažnej miere poskytuje.

1 INTRODUCTION

The fundamental concept is "agreement probability". It is shown that some undesirable cosequences of "fuzzy logic" , e.g. that tautologies of propositional calculus are not preserved can be avoided. A proposal for alternative definitions of "degree of membership" and operations on membership-graded sets is given. "Fuzziness" is interpreted as a subjectivistic concept, i.e. subjective uncertainty pertaining to the truth of a proposition. "Fuzziness" of linguistic concepts is interpreted as uncertainty of the applicability of a predicate in a given situation. This leads to the conclusion that the defnition of derived concepts, especially of hedged expressions referring to continuous scales, cannot be modelled using Zadeh´s fuzzy sets operations. Alternative interpretations of the linguistic phenomena considered and of the sorites paradox are given. Especially, the metalinguistic character of the phenomena is emphasized. Even such "nice solutions" as proposed for the sorites paradox must be rejected. It is argued, that "vagueness" and "uncertainty" should be clearly distinguished as well as "possibility" and " applicability".

Uncertainty is represented by an algorithmic mechanism built into the architecture of the solved diagnostic expert system package together with numerical belief coefficients – fuzzy sets, probabilities, and some variant of probabilities, which annotate knowledge base elements. Non-numerical belief coefficients have occasionally been used, but these do not make the underlying uncertainty semantics explicit either – words like „likely" and „unlikely" have been made to duty for probabilities or probability ranges in communicating with the user, or have been used merely as nominal annotations of propositions.

2 PROBLEM CHARACTERISTICS

A solved diagnostic expert system uses a combination of knowledge inference processes (fuzzy sets und bayesian probability) within applied domain of deliberated machinery systems.

There is, for example, illustrated more limited example of a commonly-used assumption – a „fuzzy logic" rule of combination used to entail or infer the probability of a conjunction or a disjunction from the probabilities of the conjuncts or disjuncts. Formally, it says for arbitrary propositions x_1 and x_2 (the conditional form is analogous):

$$p(x_1 \cap x_2) = \min[p(x_1).p(x_2)], \tag{1}$$

* doc. Ing., CSc., Department of Automation and Control, Faculty of Mechanical Engineering, Technical University of Košice, Park Komenského 8, Košice, tel. (+421) 55 602 2344, e-mail stefania.gallova@tuke.sk

$$p(x_1 \cup x_2) = \max[p(x_1).p(x_2)], \qquad (2)$$

This assumption was apparently used because of the perceived need to maintain unique-valued probability. In fact, it corresponds to the assumption of maximal dependency between x_1 and x_2, i.e. that either $p[(x_1)/(x_2)]$ or $p[(x_2)/(x_1)] = 1$. This is both strong and typically unjustified, as well as inconsistent in general is a rule of inference.

For example, when its use led to inconsistencies, expert system was accordingly forced to resort to renormalising probabilities. If we allow ourselves to use a bounds representation, however, we can use the following rules of inference, which are not only consistent but are sound.

They content themselves with deriving bounds on $p(x_1 \cap x_2)$ and $p(x_1) \cup x_2$:

$$\min[p(x_1).p(x_2)] \geq p(x_1 \cap x_2) \geq \max[0, p(x_1) + p(x_2) - 1],$$

$$\min[1, p(x_1) + p(x_2)] \geq p(x_1 \cup x_2) \geq \max[p(x_1), p(x_2)]$$

Thus if we have $p(x_1) = 0.2$ and $p(x_2) = 0.55$, we can use:

$$0.2 \geq p(x_1 \cap x_2) \geq 0$$

$$0.75 \geq p(x_1 \cup x_2) \geq 0.55$$

3 PROBLEM SOLVING

Further, if we assume a concrete case, that probability values come in at the lowest level of the hierarchy. The initial problem is to propagate these probabilities upward. That is given a hypothesis h_{ii}, supported by evidence – „sub-hypotheses“ $h_1, h_2, ..., h_n.$ Our idea is to compute $p(h_{jj})$. There is an assumption, that h_{jj} is at the $(m+1)$ – level of the hierarchy., and that the values $p(h_i)$ have already been accrued at the $(m\text{-}th)$ – level. H has been associated to the h_i, based on a prior model of h_{jj}, which requires $(n+k)$ component hypotheses, of which n have been observed, namely the set $(h_1, h_2, ... h_n)$. Let b_m be the number of $(m+1)$ – level hypotheses associated to each level m hypothesis. It is done from mathematical reasoning, that this mentioned value is constant at level m. There is defined a following equation:

$$p(H) = \left(\frac{1}{n+k}\right)\left(\frac{1}{b_m}\right)\sum_{i=1}^{n} p(h_i) \qquad (3)$$

We claim, that this is a probability function on any set of level $(m+1)$ hypotheses supported by evidence at level m. This fact may be mathematically proved. Suppose the probabilities p for the concrete hypotheses:

$$p(h_{11}) = 0.55$$

$$p(h_{22}) = 0.18$$

$$p(h_{33}) = 0.22$$

$$p(h_{44}) = 0.05$$

These values correspond to evidences:

$h_1 = 0.7$ *being relatively certain*

$h_2 = 0.1$ *being relatively uncertain*

$h_3 = 0.2$ *being relatively uncertain.,*

then $b_m = 2$.

From this – following a „strongest hypothesis first“ strategy – we obtain a global interpretation of (h_1, h_2, h_3, h_4). Assume without loss of generality, that all the h_i have precisely two components required in their model. Then we compute:

$p(h_{11})$ = (2/2) . (1/2) . (0.7 +1) = 0.400

$p(h_{22})$ = (2/2) . (1/2) . (0.1 + 0.2) = 0.150

$p(h_{33})$ = (1/2) . (1/2) . (0.7) = 0.175

$p(h_{44})$ = (1/2) . (1/2) . (0.2) = 0.050

These values are normalising. Given a hierarchical hypothesis space with conflicts, we invoke the rule, that all conflicts must be propagated upward. This increases the generation of alternative hypotheses, which do not claim conflicting evidence. Conflicts should be searched for at the time of hypothesis generation AND/OR associations.

All consistent interpretations of the hypothesis space should be computed and explicitly represented in the hypothesis space. This principle is illustrated in Fig. 1.

There are – alternatively – the set of possible global interpretations:

$$B = [(h_{11}, h_{44}), (h_{22,} h_{33}), (h_{33}, h_{44})] \qquad (4)$$

There is assumed, that each hypothesis has a certainty value attached which was obtained by numerically accruing evidence for each hypothesis individually, ignoring the other hypotheses. An objective is to apply a numerical accrual techniques to rank order the list of consistent interpretations. The approach suggested here is to make the list B into a probability space. There is realised a re-norming B.

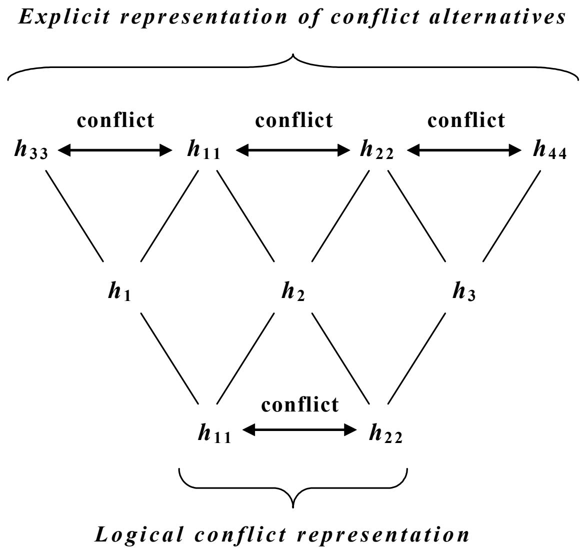

Fig. 1 Hypothesis space

Some task-dependent assumptions may have to be made. It must be – at first – derived a probability for the „joint events“ (h_{11}, h_{44}) and so on. If we assume, that – e.g. – h_{11} and h_{44} are independent, then the probability of (h_{11}, h_{44}) is simply the product of the probabilities of h_{11} and h_{44}. This seems a reasonable assumption, because if h_{11} and h_{44} are mutually consistent, then they are – probably – supported by disjoint bodies of non-contradictory evidence. We make B into a probability space by normalisation.

The list B is now rank ordered by these probabilities. From the set of $N - th$ level (conflicting) hypotheses, we can construct a set B, of mutually exclusive and exhaustic interpretations of the global situation using a straightforward recursive algorithm.

Now – is made B a probability space in the following manner. A typical element of B is a maximal set $S = (h_{11}, \ldots , h_{nn})$ of non-conflicting hypotheses. It seems reasonably to assume, that they are all mutually independent, since they are based on solved top level models and depend on disjoint bodies of evidence. Let $e = (e_1, e_2, \ldots, e_n)$, where e_i is the evidence supporting h_i.

There is:

$$p(S) = p(H_1,\ldots,H_n) = \prod_{i=1}^{n} p(H_i) \tag{5}$$

Since all the S in B are mutually exclusive and exhaustive, we can create a probability space by dividing $p(S)$ by a normalisation factor – F , calculated as:

$$F = \sum_{S \in B} P(S) \tag{6}$$

To extract the best global interpretation of the situation represented in the hierarchy, we claim it is sufficient to take the maximal consistent interpretation given by the set S with gratest probability:

$p(S)/F$, and extract the unique consistent hierarchical subspace associated below S for a hypothesis-tree. It may be proved, that this is a „probably best“ hypothesis tree in the sense that the probability that the set of hypotheses selected at level m , $(h_1, h_2, \ldots, h_n)$, by propagating the global interpretation downward from level $m + 1$, will not be the best global interpretation at level m is less than:

$$\frac{\left[\sum_{i=1}^{n} p(h_i)\right]^n}{n!} - \frac{\prod_{i=1}^{n} p(h_i)}{n!} \tag{7}$$

The methodology accounts for partial pattern matching in model-driven hypothesis generation. It was partially shown, that if a best hierarchical interpretation is inferred at the top level, then the hierarchical hypothesis tree associated to it will be optimal at all lower hierarchy levels with probability close to one. Often, the evidence and hypotheses are hierarchical in nature. In image understanding tasks, for example, evidence begins with raw imagery, from which ambiguous features are extracted which have multiple possible aggregations providing evidential support for the presence of multiple hypotheses of objects and terrain, which in turn aggregate in multiple ways to provide partial evidence for different interpretations of the ambient scene.

Information fusion for manufacturing situation understanding has a similar „evidence – hypothesis“ hierarchy from multiple sensor through message level interpretations, and also provide evidence at multiple levels of the machinery diagnostic system hierarchy.

The probabilistic learning system can be used for „single concept“ learning, but most testing has been in the more complex domain of heuristic search. Here noise arises because of inadequate features, search anomalies, and changing environments.

A critical question in this discussion is whether we can represent all of the information we need to solve our problems through a suitable probability distribution.

The first example will be part of the problem of learning algebraic notation from examples. Let the examples are of the following form:

33, 49, + : 82

4, 9, x : 36

-7, 3, + : -4 *and so on.*

The examples all use +, -, x, and – only. The problem is for the machine to induce the relationship of the object to the right of the colon, to the rest of the expression. To do this, it has a vocabulary of seven symbols, which are:

I_1, I_2, I_3 represent the first three symbols of its inputs

For example – „33“ „49“, and +.

Add, Sub, Mul, Div, ... represent internal operators, that can operate on the contents of I_1, I_2, I_3, if they are numbers.

The system tries to find a short sequence of these seven symbols, that represents a program expressing the symbol to the right of the colon in terms of the other symbols:

33, 49, +, : 82 can be rewritten as 33, 49, + : I_1, I_2, *Add*.

If all symbols have equal probability to start, the subsequence I_1, I_2, *Add* has probability:

1/7 x 1/7 x 1/7 = 1 / 343

Then, if we assume 16 bit precision, each integer, such as 82, has probability 2^{-16}=1 / 65536.

So 1 / 343 is a great improvement over the original data.

In the following, we can code the right halves of the original expressions as:

I_1 , I_2 , *Add*

I_1 , I_2 , *Mul*

I_1 , I_2 , *Add*

The probability, that the symbol in the first position is I_1 , is close to one (similarly the second symbol). This gives a probability of close to 1/7 for I_1, I_2, *Add*. If we can increase the probability of solved code further by noting, that in the expression like 33, 49, + : I_1, I_2, *Add*, the last symbol is closely correlated with the third symbol. So that knowing the third symbol, we can assign very high probability to the final symbols that actually occur. If we use many parallel codes for a symbol, we can have an equivalent code length of exactly, - log *p*. A probability of close to *1* corresponds to an equivalent code length of close to zero.

4 CONCLUSIONS

The methodology accounts for partial pattern matching in model-driven hypothesis generation. This designed system shows, that if a best hierarchical interpretation is inferred at the top level, then the hierarchical hypothesis tree associated to it will be optimal at a lower hierarchy levels with probability close to one.

When using the diagnostic expert system parallel combination function, one is implicitly assuming this form of conditional independence. There was examined a consequence of this assumption and its combination with real fuzzy reasoning within this expert system, which demonstrates their practical application.

In this system, there are several restrictions, that a system must satisfy before the use of fuzzy and probabilistic combination reasoning within inference diagnostic process. For example, the system contains a set of mutually exclusive and exhaustive hypotheses with more than two elements and its inference network is not a tree.

Future research is necessary to determine the practical value of the present approach compared with other approaches for using contingency tables for the other uncertain inference. Additional effort might further improve the computational efficiency of the present method for use in a full scale implementation.

These results are based on the topology of the „influence diagram“ graph, and do not depend upon the actual sample spaces or probability distributions. Since arcs may be present in the graph even when random variables are conditionally independent, these results may overstate the need for information. It is therefore important to capture a natural influence diagram in the first place, which would tend to show considerable conditional independence.

The description of the problem and the context are entered by the user or by the system. There are various ways of communicating with the system – by natural, graphical or problem oriented language. A problem in this connection is the real-time capability of knowledge-based (expert) systems. Application involving sensors, usually, must have a control system which can take actions in real time.

There is a manufacturing knowledge module, which can be understood as a world model of the domain for which the expert system was developed. It is like a huge database which contains all factual knowledge and rules needed for the operation of the expert system. There are also empirical rules in the form of meta-knowledge stored in the database, which knows how the factual knowledge has to be processed to find an answer to an inquiry. There are two ways of operating this module.

First, the user enters into the system the description (context) of his problem. It is stored for use in the system. Second, the system constructs the description by interrogating the user in a question and answer session. For this operation, there must be a description of the problem available to the system. This module can be considered as a temporary storage needed for the solution of the problem.

Inference engine module is the knowledge processor, which looks at the problem description and tries to find a solution with the help of both, the factual and meta-knowledge. The user can communicate with the explainer to obtain a report about the operation of the expert system.

Expert system for diagnosis is the most advanced artificial intelligence tool used in manufacturing. It plays an important role in supervising complex production equipment and locating problems as soon as they arise. It is possible to reduce the average search time for a detective machine component from an hour to a few minutes.

This problem is solved in terms of grant project VEGA: "Diagnostic System Building of Machinery Equipment with Knowledge-Based System and Chaos Theory Applying" (Number: VEGA 1/1084/04).

REFERENCES

[1] KOVÁČ, D. & KOVÁČOVÁ, I. Behavior of the Voltage Measurement Transformer during Non Harmonic Signal Measuring. In *Power Electronics and Motion Control*. Czech Republic : 1998, pp. 152-155. ISBN 80-01-01832-6.

[2] REMBOLD, U., BLUME, C. & DILLMANN, R. *Computer-Integrated Manufacturing Technology and Systems.* New York : Marcel Dekker New York, 1995. 250 pp. ISBN D-201-56512-1.

[3] RIMÁR, M. The Stabilization of Low Damped Astatic Systems by Means of Adaptive Proportional Time-Delay Filter. *Transactions of the Universities of Košice*. 2004, Nr. 1, pp. 17-24. ISSN 1335-2334.

[4] SALOKY, T., PITEĽ, J. & VOJTKO, I. Robot Safety Evaluation. In *Automatizácia – Robotika v teórii a praxi ROBTEP 99*. Prešov : 1999, pp. 257-260. ISBN 80-7099-388-X.

Reviewer: doc. Ing. Miluše Vítečková, CSc., VŠB-Technical University of Ostrava