University of Ostrava

Institute for Research and Applications of Fuzzy Modeling

Preselection of Rules in Fuzzy Logic Deduction

Antonín Dvořák

Research report No. 17

January 15, 1999

Submitted/to appear:

International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems

Supported by:

201/96/0985 GA AV ČR

University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
Bráfova 7, 701 03 Ostrava 1, Czech Republic

tel.: +420-69-622 2808 fax: +420-69-22 28 28
e-mail: antonin.dvorak@osu.cz

International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems
Vol. 0, No. 0 (1993) 000—000

(Leave $1\frac{1}{2}$ inch blank space for Publisher)

ON PRESELECTION OF RULES IN FUZZY LOGIC DEDUCTION

ANTONÍN DVOŘÁK
Institute for Research and Applications of Fuzzy Modeling
Bráfova 7, 701 03 Ostrava, Czech Republic
e-mail: antonin.dvorak@osu.cz

Received (received date)
Revised (revised date)

Preselection of rules is an algorithm which chooses one rule from linguistic description of some process. This rule is then used for inference instead of the whole description. This article studies properties of preselection and proposes use of measures of resemblance for choosing the rule which best fits the observation.

Keywords: Preselection, Fuzzy inference, Approximate reasoning, Similarity measures.

1. Introduction

Fuzzy logic and fuzzy control became very popular recently. However, there exist several problems when applying fuzzy logic to real world situations. One of the most difficult ones is the *computational complexity* of fuzzy inference mechanisms[1,2]. In the previous paper[2] were presented three methods how to deal with the computational complexity problem. Further experiments and investigations show that the best results can be obtained by means of the method based on preselection of the most suitable rule from linguistic description. Theoretical and practical aspects of such preselection algorithm will be studied in this paper.

The basic scheme for inferring conclusions from observations when linguistic description (of some physical process, decision-making situation etc.) is given, known as *generalized modus ponens*, is the following:

$$\begin{array}{rl} \text{Condition:} & \mathcal{R}_1 := \mathsf{IF}\ X_1\ \mathsf{is}\ \mathcal{A}_{11}\ \mathsf{AND} \ldots\ \mathsf{AND}\ X_n\ \mathsf{is}\ \mathcal{A}_{n1}\ \mathsf{THEN}\ Y\ \mathsf{is}\ \mathcal{B}_1 \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & \mathcal{R}_r := \mathsf{IF}\ X_1\ \mathsf{is}\ \mathcal{A}_{1r}\ \mathsf{AND} \ldots\ \mathsf{AND}\ X_n\ \mathsf{is}\ \mathcal{A}_{nr}\ \mathsf{THEN}\ Y\ \mathsf{is}\ \mathcal{B}_r \\ \text{Observation:} & X_1\ \mathsf{is}\ \mathcal{A}'_1\ \mathsf{AND} \ldots\ \mathsf{AND}\ X_n\ \mathsf{is}\ \mathcal{A}'_n \\ \hline \text{Conclusion:} & Y\ \mathsf{is}\ \mathcal{B}', \end{array}$$

where $X_1, \dots, X_n$ are the *antecedent variables*, Y is the *succedent variable*, and $\mathcal{A}'_1, \dots, \mathcal{A}'_n$ are expressions of natural language, which may be slightly different from $\mathcal{A}_{1j}, \dots, \mathcal{A}_{nj}$. Hence, the conclusion $\mathcal{B}'$ can be slightly different from all $\mathcal{B}_j,\ j = 1, \dots, r$. The general form of the considered linguistic expressions is

$$[\langle\text{linguistic modifier}\rangle]\langle\text{atomic term}\rangle.$$

Here, ⟨linguistic modifier⟩ is an intensifying adverb with narrowing or extending effect (e.g. *very, more or less* etc.) and the negation *not*. The ⟨atomic term⟩ may be an adjective (we use *small, medium, big*), a fuzzy number or some special term (e.g. *undefined*). A typical example of the linguistic expression $\mathcal{A}$ is *very small*, *roughly big* etc.

The fundamental formula which describes inference mechanism based on the above schema is

$$B'y = \bigvee_{x_1 \in U_1, \dots, x_n \in U_n} ((A'_1 x_1 \wedge \dots \wedge A'_n x_n) \otimes \bigwedge_{j=1}^{r} ((A_{1j} x_1 \wedge \dots \wedge A_{nj} x_n) \to B_j y)), \tag{1}$$

where $U_1, \dots, U_n$ are universes of discourse of antecedent variables, $A_{1j}, \dots, A_{nj}$, $A'_1, \dots, A'_n$, B_j are fuzzy sets which represent the meanings of linguistic expressions in the antecedent part of the linguistic description, observations and linguistic terms in the consequent part of linguistic description, respectively. Construction of these meanings is described in the paper[3]. $\otimes$ and $\to$ are the Łukasiewicz conjunction and implication, respectively. For details, see e.g.[4] In the following, we will call the above described inference mechanism as *fuzzy logic deduction*.

The most important distinction between fuzzy logic deduction and so-called Max-Min fuzzy inference[5] is the following. Fuzzy logic deduction is based on many-valued Łukasiewicz logic. If-THEN rules from the linguistic description are understood as linguistically expressed logical implications. The derivation of formula (1) respects it. On the other hand, Max–Min inference is in fact interpolation of unknown partially given function by means of fuzzy graph[6]. There is no implication or logical modus ponens inference rule behind.

2. Computational complexity

When we analyze the behavior of our inference mechanism, we have to distinguish three situations:

- observations are *crisp numbers*,
- observations are *fuzzy singletons*,
- observations are *fuzzy sets*.

In the first case we are able to simplify (1) to

$$B'y = \bigwedge_{j=1}^{r} ((A_{1j}x_1 \wedge \cdots \wedge A_{nj}x_n) \to B_jy), \tag{2}$$

where $x_1, \cdots, x_n$ are crisp observations.

If we denote $\gamma = (A'_1x_1 \wedge \cdots \wedge A'_nx_n)$, where $A'_1, \ldots, A'_n$ are fuzzy singletons, then the second case leads to formula

$$B'y = \gamma \otimes \bigwedge_{j=1}^{r} ((A_{1j}x_1 \wedge \cdots \wedge A_{nj}x_n) \to B_jy). \tag{3}$$

For practical reasons, we assume that the universes of discourse of all the variables are discrete, i.e.

$$U_i = \{x_1, x_2, \ldots, x_P\}, \quad i = 1, \ldots, n,$$

$$V = \{y_1, y_2, \ldots, y_P\},$$

where the number P of elements is the same for all $U_i, i = 1, \ldots, n$ and V.

The number of arithmetical operations required to computation of conclusion B' is in these cases asymptotically equal to

$$\mathcal{C}_{1,2} = \mathcal{O}(r(n+1)P). \tag{4}$$

In the third case, i.e. when $A'_1, \ldots, A'_n$ are fuzzy sets, the computational complexity can by characterized as

$$\mathcal{C}_3 = \mathcal{O}(rP^{n+1}). \tag{5}$$

Note that r in (5) is also a function of n, so that the complexity is in fact $\mathcal{O}(P^{2n+1})$. It is still an open question whether it is possible to avoid the exponentiality in (5). In paper[2] we propose several methods which can be used to speed up computation, but none of them allows to get rid of the exponentiality. The shape of (5) causes the inference with fuzzy observations to be practically intractable for $n \geq 4$.

One of rather natural ways for speeding inference up is called *preselection of rules*. The idea is simple: we don't use the whole linguistic description, but choose in advance only one rule and perform the inference with it. But there arise several questions connected with this idea, namely:

- what rule should be chosen?
- how it influences the behavior of the inference?

These questions we attempt to answer in the foregoing sections. We will describe the hardest case, i.e. fuzzy logic deduction with fuzzy observations.

3. Preselection

We can regard the problem of choosing the right rule from several viewpoints. We can try to find the rule which gives the same or at least most similar result to the situation when all linguistic description is involved. On the other hand, we may aim to achieve even better results in some situations when results produced by the inference with the whole linguistic description are contraintuitive.

As example let us consider the following linguistic description:

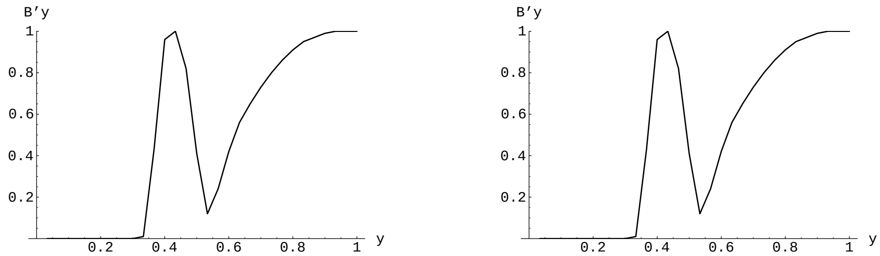
IF X is *small* THEN Y is *big*
IF X is *very small* THEN Y is *medium*

The following figures show how this in our opinion quite reasonable description behave with respect to input fuzzy sets A' which represent the meanings of *small* and *very small*. These figures show that Zadeh–Mamdani inference mechanism can't distinguish between these two inputs at all (Figure 1). We see that the inferred fuzzy sets are the same for both observations. The results obtained by means of logical deduction shows that this method can somehow discriminate among inputs, but the shape of resulting fuzzy sets is not satisfactory (Figure 2). The subnormality of resulting fuzzy sets is typical for fuzzy logic deduction when some rules from linguistic description are in conflict, i.e. there are rules with similar antecedents (*small* and *very small*) and different consequents. But there are situations when it is natural to use such a rules. For example, let X in the above description be a price of some commodity, and let Y be our preference to buy it. Low price is of course important, but too cheap things are suspicious. We see that Max–Min inference as well as logical deduction without preselection are not able to distinguish well between these situation. This problem can be in the frame of fuzzy logic deduction solved by preselection of rules. When preselection algorithm was plugged in, then the results has been exactly as we intuitively expected (Figure 3), i.e. 'Big' when observation was *small* and 'Medium' when observation was *very small*.

Theoretical background behind preselection of rules can be described in the frame of the theory of approximate reasoning[7,4]. Let us consider generalized modus ponens scheme from Section 1 (for the sake of simplicity, let there be one antecedent variable). Then we can apply the following inference rule of approximate reasoning: (Modus ponens with conjunction of implications):

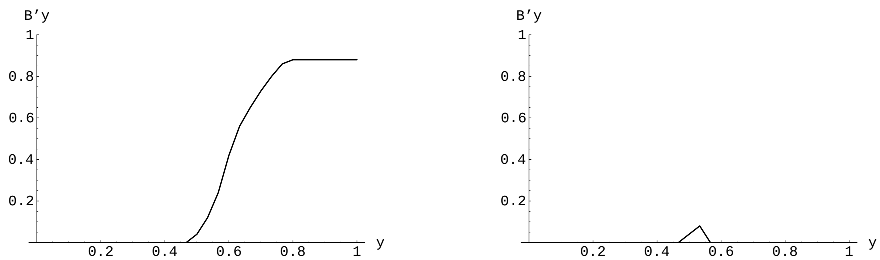
$$\frac{\mathcal{A}_k,\ \text{IF } \mathcal{A}_1 \text{ THEN } \mathcal{B}_1 \text{ AND} \ldots \text{ AND IF } \mathcal{A}_r \text{ THEN } \mathcal{B}_r}{\mathcal{B}_k} \quad \left[\frac{\underline{A_k}, \bigwedge_{j=1}^{r} \underline{A_j \Rightarrow B_j}}{\underline{B_k}}\right],$$

for $1 \le k \le r$, where $\underline{A_k}, \underline{A_j \Rightarrow B_j}, \underline{B_k}$ are *multiformulas*[7], i.e. sets of evaluated formulas which express vague properties. Let us suppose that for every individual observation $\mathcal{A}$ there is k-th rule which describes action that should be taken. We only don't know, which rule from linguistic description it is. Our task then



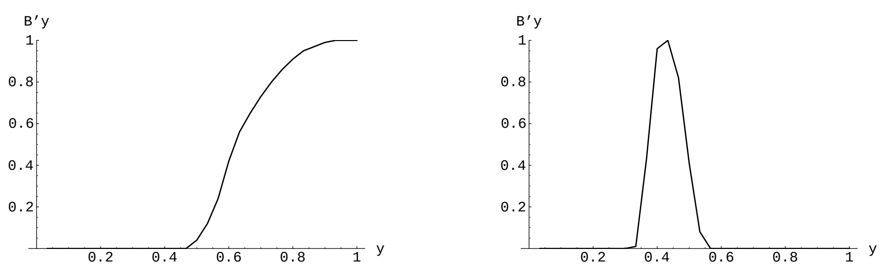
(a) $A' =$ Small (b) $A' =$ Very Small

Figure 1: Max-min inference



(a) $A' =$ Small (b) $A' =$ Very Small

Figure 2: Logical deduction without preselection



(a) $A' =$ Small (b) $A' =$ Very Small

Figure 3: Logical deduction with preselection

can be understood as a determination of this rule. If we find out that observation is exactly equal to one of the antecedents, then we can finish the searching. If it is not the case, we have to try to find the rule which antecedent is most similar to our observation. As soon as this rule is found, we use it for inferring the conclusion, i.e. instead of the whole description $\mathcal{R} = \{\mathcal{R}_1, \ldots, \mathcal{R}_k, \ldots \mathcal{R}_r\}$ we use $\mathcal{R}' = \{\mathcal{R}_k\}$. We propose to use a *similarity measure* to determine that rule.[1]

4. Similarity measures

In order to compare two fuzzy sets, e.g. observation and antecedent part of IF–THEN rule, we need some mapping which can measure their similarity[8]. However, not all similarity measures described there are suitable for use in the preselection algorithms. The problem is that we are interested in influence which observation will take in inference process. It is clear, that even very dissimilar sets can "match" in such a way that appropriate rule fires.

Here we propose to apply approach described in[9], which is an attempt to make a framework for the measures of comparison of objects.

Definition 1 A fuzzy set measure M *is a mapping:* $F(\Omega) \to \mathbb{R}^+$ *such that, for every* A *and* B *in* $F(\Omega)$*:*

1. $M(\emptyset) = 0$,
2. $B \subseteq A \Rightarrow M(B) \leq M(A)$.

Definition 2 M-measure of comparison *on* Ω *is a mapping* $S : F(\Omega) \times F(\Omega) \to [0, 1]$ *such that*

$$S(A, B) = F_S(M(A \cap B), M(B - A), M(A - B))$$

for a given mapping

$$F_S : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \to [0, 1]$$

and a fuzzy set measure M *on* Ω.

M-measures of comparison are in[9] further divided to the following categories: measures of satisfiability, measures of inclusion and measures of resemblance. From these types we consider as the best for our purpose measures of resemblance, which are defined as follows:

Definition 3 A M-measure of resemblance *on* Ω *is an* M*-measure of comparison* S *which satisfies*

1. $F_S(u, v, w)$ *is non-decreasing in* u*, non-increasing in* v *and* w*,*
2. $S(A, A) = 1$,

[1]Note that similar preselection mechanism can be used also with Max–Min inference rule. However, it seems to be hard to find the theoretical justification of it.

3. $S(A,B) = S(B,A)$.

Examples of M-measure of resemblance are

$$S_1(A,B) = \frac{M(A \cap B)}{M(A \cap B) + M(A \triangle B)}$$

$$S_2(A,B) = 1 - \frac{M(A \triangle B)}{M(A) + M(B)}$$

where

$$M(A) = \sum_{x \in \Omega} Ax,$$

$$(B - A)x = \max(0, Bx - Ax),$$

$$(A \triangle B)x = |Ax - Bx|.$$

An example of M-measure of similitude, which is not M-measure of resemblance [8]:

$$S_3(A,B) = \frac{M(A \cap B)}{M(A \cap B) + \alpha M(B - A) + \beta M(A - B)}, \quad \alpha, \beta \in [0,1].$$

These measures satisfy additional property

$$S(A,B) = 1 \Leftrightarrow A = B,$$

which is called *separability*.

In our algorithm we use these measures of resemblance, which are separable. These measures has, in our opinion, all properties necessary for use in the preselection algorithm.

5. Preselection algorithm

In this section we present an algorithm of preselection of rules.
The algorithm of computation of B' is:

1. For all rules $\mathcal{R}_j,\ j = 1, \ldots, r$ we determine overall similarity measure S_j by
$$S_j = \prod_{i=1}^{n} S(A_i, A_i'). \tag{6}$$
2. Choose rule $\mathcal{R}_l$ with the highest S_l.
3. Compute the conclusion B' by means of
$$B'y = \bigvee_{x_1 \in U_1, \ldots, x_n \in U_n} ((A_1' x_1 \wedge \cdots \wedge A_n' x_n) \otimes ((A_{1l} x_1 \wedge \cdots \wedge A_{nl} x_n) \to B_l y)). \tag{7}$$

If there occurs (quite unlikely) situation, when two (or more) rules have the same S_j, then we propose to use as auxiliary criterion the sum of Hausdorff distances[8] between the kernels of (normalized, if necessary) fuzzy sets A_i, A'_i, $i = 1, \dots, n$. If we denote Hausdorff distance between kernels of A_i, A'_i as $H(A_i, A'_i)$, then we choose rule with the smallest

$$H_j = \sum_{i=1}^{n} H(A_i, A'_i).$$

The usage of Hausdorff distance of kernels is motivated by endeavour to stress the importance of α-cuts with the highest membership degrees.

The reason for using product in (6) is that we can imagine inference with several antecedent variables as follows: Fuzzy sets A_i, A'_j, $i, j = 1 \dots n$ are orthogonal projections of multidimensional fuzzy sets $\mathbf{A}$, $\mathbf{A}'$. Then, it is generally not possible to reconstruct $\mathbf{A}, \mathbf{A}'$ from A_i, A'_j. Nevertheless, we can found an upper approximation of "volumes" of $\mathbf{A}, \mathbf{A}'$ by computing product of "areas" of A_i, A'_j. The same argument applies for the computation of the measure of resemblance.

The speeding-up factor is approximately equal to the number of rules in the linguistic description r. It approaches r when the number of antecedent variables increases, because the part of computations which pertains to the determination of similarity measures becomes negligible.

6. Conclusion

We presented an algorithm for preselection of rules in the frame of fuzzy logic deduction. We claim that this algorithm speeds-up the computation of conclusions, as well as it improves the performance of deduction in situations when there are more rules which match observation approximately equally. One of advantages of this algorithm is that we are not forced to use any threshold for the similarity measure, because if the maximal similarity is small, we can expect that as a result of inference we obtain highly "nonspecific" fuzzy set, i.e. fuzzy set which membership function has value close to 1 for all members of universe of the consequent variable.

Similar approach, namely *analogical reasoning*, has been proposed in[10] (see also[11]). It uses also similarity measure for determination of the best fitting rule. Conclusion is then obtained by means of some *modification function* on membership function of fuzzy set which models the meaning of consequent linguistic expression of selected rule, provided that similarity measure is greater than some threshold selected in advance. Main points in which both approaches differ are the following: first, we have not to think about modification functions depending on similarity between observation and premises, because we use formula (7) for computation of conclusions. Using this formula, conclusion will behave according to: "the more is observation similar to antecedent part of the chosen rule, the more should be the conclusion similar to consequent part". Second, for similar reasons we don't need to determine any threshold for similarity.

Our further research will be focused to detailed study of behavior of similarity measures used in preselection algorithm. It can be also interesting to examine relationship of this algorithm to usage of gradual rules which were studied by D. Dubois and H. Prade[12].

Acknowledgement

This research was supported by the grant 201/96/0985 of the GA ČR.

References

1. L. T. Kóczy: "Algorithmic Aspects of Fuzzy Control", *Int. J. of Approximate Reasoning* **12** (1995), 159–219.
2. A. Dvořák: Computational Properties of Fuzzy Logic Deduction. In: *Computational Intelligence. Theory and Applications. Proceedings of the 5th Fuzzy Days Dortmund*, ed. B. Reusch (Springer–Verlag, Berlin-Heidelberg 1997), 189–195.
3. V. Novák and I. Perfilieva: "Evaluating Linguistic Expressions and Functional Fuzzy Theories in Fuzzy Logic", to appear in: *Computing with Words in Systems Analysis.* eds. L. A. Zadeh and J. Kacpryk (Springer-Verlag, Heidelberg, 1999).
4. V. Novák: "Linguistically Oriented Fuzzy Logic Controller and Its Design", *Int. J. of Approximate Reasoning* **12**(1995), 263–277.
5. G. J. Klir and Bo Yuan: *Fuzzy Sets and Fuzzy Logic. Theory and Applications* (Prentice Hall, Englewood Cliffs 1995).
6. F. Klawonn and V. Novák: "The Relation between Inference and Interpolation in the Framework of Fuzzy Systems", *Fuzzy Sets and Systems* **81**(1996), 331–354.
7. V. Novák, I. Perfilieva and J. Močkoř: *Mathematical Principles of Fuzzy Logic* (Kluwer, Dordrecht, to appear).
8. R. Zwick, E. Carlstein and D. V. Budescu: "Measures of Similarity Among Fuzzy Concepts: A Comparative Analysis", *Int. J. of Approximate Reasoning* **1**(1987), 221–242.
9. B. Bouchon-Meunier, M. Rifqi and S. Bothorel: "Towards General Measures of Comparison of Objects", *Fuzzy Sets and Systems* **84**(1996), 143–153.
10. I. B. Turksen and Z. Zhong: "An Approximate Analogical Reasoning Schema Based on Similarity Measures and Interval Valued Fuzzy Sets", *Fuzzy Sets and Systems* **34**(1990), 323–346.
11. E. S. Lee and Q. Zhu: *Fuzzy and Evidence Reasoning* (Physica–Verlag, Heidelberg 1995).
12. D. Dubois and H. Prade: "What are fuzzy rules and how to use them", *Fuzzy Sets and Systems* **84**(1996), 169–185.
13. V. Novák: *Fuzzy Sets and Their Applications* (Adam Hilger, Bristol 1989).
14. L. A. Zadeh: "Outline of a New Approach to the Analysis and Synthesis of Complex Systems and Decision Processes", *IEEE Trans. System Man Cybernet.*, **1**(1973), 28–44.