8-1-2023

If Everything is a Matter of Degree, Why Do Crisp Techniques Often Work Better?

Miroslav Svitek
Czech Technical University in Prague, svitek@fd.cvut.cz

Olga Kosheleva
The University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Comments:

Technical Report: UTEP-CS-23-51

If Everything is a Matter of Degree, Why Do Crisp Techniques Often Work Better?

Miroslav Svítek, Olga Kosheleva, and Vladik Kreinovich

Abstract Numerous examples from different application domain confirm the statement of Lotfi Zadeh – that everything is a matter of degree. Because of this, one would expect that in most – if not all – practical situations taking these degrees into account would lead to more effective control, more effective prediction, etc. In practice, while in many cases, this indeed happens, in many other cases, "crisp" methods – methods that do not take these degrees into account – work better. In this paper, we provide two possible explanations for this discrepancy: an objective one – explaining that the optimal (best-fit) model is indeed often the crisp one, and a subjective one – that we have to use crisp because of our limited ability to process information.

1 Formulation of the Problem

On the one hand, everything is a matter of degree. In many application areas, our knowledge and expertise is described by using imprecise ("fuzzy") words from natural language. Financial experts talk about *significant* decreases or increases of stock prices, medical doctors talk about *high* fever and *irregular* shape of skin formations, skilled drivers talk about the danger of cars being *too close*, etc. None of these terns is precisely defined, and this makes sense – if they were precisely defined, we would not need these experts, we could simply follow the well-defined rules.

Lotfi Zadeh, the father of fuzzy techniques – that enable us to transform such imprecise knowledge into computer-understandable mathematical form – formulated

Miroslav Svítek
Faculty of Transportation Sciences, Czech Technical University in Prague, Konviktska 20
CZ-110 00 Prague 1, Czech Republic, e-mail: svitek@fd.cvut.cz

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso, 500 W. University, El Paso, Texas 79968, USA
e-mail: olgak@utep.edu, vladik@utep.edu

this as a mantra that he likes to repeat in this talks: *Everything is a matter of degree.* This mantra was the foundation of his techniques, where, to describe an imprecise term like "small" in precise terms, we assign:

- to each possible value x of the corresponding quantity,
- the degree $\mu(x)$ from the interval $[0,1]$ to which,in the expert's opinion, this value satisfies the given property: e.g., to what extent x is small.

The function $\mu(x)$ is called a *membership function*, or a *fuzzy set*. Here:

- the degree 1 means that this value definitely satisfies this property,
- the degree 0 means that this value definitely *does not* satisfy this property, and
- values between 0 and 1 correspond to the same when this property is satisfied to some degree.

In this description, degrees 0 and 1 correspond to the fuzzy case, while degree intermediate between 0 and 1 correspond to the truly precise ("crisp") case.

This mantra has been very successful. Zadeh's general idea had led to many successful applications of fuzzy techniques and fuzzy ideas; see, e.g., [1, 4, 5, 7, 8, 11].

But why not more applications? One would expect that with this natural idea, the majority of applications would start using some version of fuzzy techniques and gain even more successes – but this has not happened.

For example, while fuzzy control – the most well-known applications of fuzzy techniques – had many impressive successes, the majority of control systems are "crisp"– they do not take into account imprecise knowledge, and they therefore do not use fuzzy techniques.

Could the reason be sociological? Maybe the problem is that researchers and practitioners are too conservative, they stick to their traditional methods, they do not understand how fuzzy techniques can be helpful – and it is our goal to promote these techniques?

This sociological argument may have been somewhat convincing in the past, but nowadays, when everyone use neural techniques – which were, in the past, as unappreciated as fuzzy one – this is no longer a convincing argument.

Probably in many cases, crisp methods work better. Why is everyone is using neural techniques? Because in many cases, they have been successful; see, e.g., [2]. Why not everyone uses fuzzy techniques? A natural neural-motivated answer is: because in many cases, crisp methods work better.

A simple example is current machine learning techniques like deep learning. In a few cases, researchers added fuzzy information and got better results, but in many other cases, traditional (crisp) neural networks work so well that adding fuzzy only decreases their effectiveness.

Resulting puzzle. There is a clear discrepancy:

- On the one hand, there are convincing arguments that everything is a matter of degree – which should imply that taking these degrees into account should improve the effectiveness of data processing, control, predictions, etc.
- On the other hand, in many practical situations, crisp techniques – that do not take these degrees into account – work better.

How can we explain this discrepancy?

What we do in this paper. In this paper, we provide two explanations:

- an objective one – that crisp methods are indeed often better, and
- a subjective one – that due to our limited perception abilities, we perceive many situations as crisp, even though in reality they are somewhat fuzzy.

2 An Objective Explanation

It is all optimization. How do we come up with a model of reality? There are usually several possible models, and we select the one that provides the *best* fit with observations and measurements.

In mathematical terms, "the best" means optimal. So, whatever model we use is the result of optimization.

What do we know about optimization? Optimization is one of the main applications of calculus – and one of the main reasons why calculus was invented in the first place.

According to calculus, in the ideal situations, when there are no constraints, the maximum or minimum of a function $f(x_1, \ldots, x_n)$ is attained at a point where all its partial derivatives are zeros:

$$\frac{\partial f}{\partial x_1} = \ldots = \frac{\partial f}{\partial x_n} = 0.$$

There are usually several such "stationary" points: some of them describe the desired global maximum or minimum, some of them describe local maxima and minima, some are saddle points – which are neither maxima not minima.

In practice, there are always many constraints. For example, the need to fit the observations very strongly restricts possible models. As we gather more and more constraints, the set S of possible solutions – that satisfy all the constraints – becomes smaller and smaller. As a result, in general, this set no longer contains any stationary points.

In this case, according to calculus, the optimal solution to the corresponding constrained optimization problem is attained at the *boundary* of the set S of possible solutions. Indeed, if the maximum or minimum were obtained in the interior of this set, then, according to calculus, all partial derivatives of the objective function at this point would be equal to 0, so this would be of the stationary points of the objective function – and we assumed that the set S no longer contains any stationary points.

Let us apply this idea to our puzzle. We want to find the values of the degree d_i that provide the best description of the observed phenomena. The set of possible values of each degree d_i is the interval $[0, 1]$. Thus, the boundary of this set consists of two points – 0 and 1.

So, a simple application of calculus implies that in most case, the most adequate model is the one in which all the degree are either 0s or 1s – i.e., a crisp model.

This explains why in many case, crisp techniques work better.

Comment. It should be mentioned that this argument is not completely new: we had a similar argument – for a somewhat different purpose – in [10].

This is not just a theoretical conclusion. One may think that our conclusion is too abstract and too theoretical to be convincing, but this conclusion is confirmed by several more down-to-earth examples.

For example, in [3], the authors assumed that the membership functions are trapezoidal, and used optimization techniques to find the parameters of these functions that lead to the best results. Interestingly (and unexpectedly), the optimal membership functions turned out to be crisp ones, where the membership function is always equal either to 0 or ro 1.

3 A Subjective Explanation

Let us now consider various models – fuzzy or crisp – from the subjective viewpoint.

How subjective knowledge is usually described. Usually, one knowledge about the control or prediction is described by rules. Namely, if we are interested in predicting (or determining) the value of the desired quantity q based on the values of the known quantities $q_1, \ldots, q_m$, then we do the following:

- for each of m inputs and for the desired output, we select several levels – e.g., small, medium, large, somewhat large, etc.;
- for each combination of input levels, we described the expected level of y.

For example, if we have two levels for each q_i – small and large – and two variables ($m = 2$), then we can have rules like the following:

if q_1 is small and q_2 is small, then y is small,
if q_1 is small and q_2 is large, then y is medium, etc.

Our abilities are limited: general fact. It is known – see, e.g., [6, 9] – that we can, in general, process no more than seven plus minus two objects in our active mind. Some people can only process $7 - 2 = 5$ objects, some can process $7 + 2 = 9$, but practically no one can process more.

Let us apply this general fact to our case. From the viewpoint of the above-described seven-plus-minus-two law, it makes sense to conclude that to be able to meaningfully use the system of rule, we must have no more than seven plus minus two rules in this sequence:

- at most five rules if we want *all* people to be able to use these rules, and
- at most nine rules if we want at least *some* people to be able to use these rules.

What does this imply about the number of levels? For each input, we should have at least two levels: otherwise, the conclusion does not depend on this input at all.

How many rules can we have? Let L_i be the number of levels selected for the i-th quantity q_i. Then, we have as many rules as there are combinations of these levels – i.e., $R = L_1 \cdot \ldots \cdot L_m$ rules.

If we have only one input q_1, then we have as many rules as there are levels: $R = L_1$. Thus, the above restriction on the number of rules means that we can have no more than seven plus minus two levels – and this is what most applications of fuzzy techniques do.

If we have two inputs q_1 and q_2, then ideally, we should have $L_1 \cdot L_2 \le 5$ rules. Since each number of levels L_i is at least two, this means that in this case, we need to have exactly two levels for each input.

If we want the rules to be usable by *some* people, then we should have $L_1 \cdot L_2 \le 9$. Here, we have several choices for the pair (L_1, L_2):

$$(2, 2), (2, 3), (3, 2), (3, 3), (2, 4), (4, 2).$$

If we have three inputs, then there is no way to make rules usable by everyone. If we want to have rules that can be used by some people, then we must have $L_1 \cdot L_2 \cdot L_2 \le 9$. Under the condition $L_i \ge 2$, there is only one way to satisfy this inequality: namely, to have $L_1 = L_2 = L_3 = 2$.

This is exactly what we wanted to explain. If we have only two levels – e.g., small or large – this means that we ignore all the nuances and, in effect, consider a crisp case: either small or not small, either large or not large.

This is what we meant by subjective reason: we may, in principle, have many possible degrees, but the limitations of our perception and information processing ability leads to the use of crisp objects.

Comment. If we have more than 3 inputs, we cannot have no more than 9 rules: even if for each input we have two different levels, with at least 4 inputs we will have at least $2^4 = 16$ rules. Thus, our intuitive rules are limited to the cases of no more than three inputs.

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in

Computer Science), HRD-1834620 and HRD-2034030 (CAHSI Includes), EAR-2225395 (Center for Collective Impact in Earthquake Science C-CIES), and by the AT&T Fellowship in Information Technology.

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

The authors are thankful to all the participants of the 20th World Congress of the International Fuzzy Systems Association IFSA'2023 (Daegu, South Korea, August 20–23, 2023) for valuable discussions.

References

1. R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.
2. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, Massachusetts, 2016.
3. H. Irie and I. Hayashi, "Evaluation of bagging-type ensemble method generating virtual data", *Proceedings of the 20th World Congress of the International Fuzzy Systems Association IFSA 2023*, Daegu, South Korea, August 20–23, 2023, pp. 282–288.
4. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
5. J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
6. G. A. Miller, "The magical number seven plus or minus two: some limits on our capacity for processing information", *Psychological Review*, 1956, Vol. 63, No. 2, pp. 81–97.
7. H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.
8. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
9. S. K. Reed, *Cognition: Theories and Application*, SAGE Publications, Thousand Oaks, California, 2022.
10. M. Svitek, O. Kosheleva, and V. Kreinovich, "Is fully explainable AI even possible: fuzzy-based analysis", *Proceedings of the 20th World Congress of the International Fuzzy Systems Association IFSA'2023*, Daegu, South Korea, August 20–23, 2023, pp. 259–264.
11. L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.