

Why Fuzzy Techniques in Explainable AI?

Which Fuzzy Techniques in Explainable AI?

Kelly Cohen, Laxman Bokati, Martine Ceberio, Olga Kosheleva, and
Vladik Kreinovich

Abstract One of big challenges of many state-of-the-art AI techniques such as deep learning is that their results do not come with any explanations – and, taking into account that some of the resulting conclusions and recommendations are far from optimal, it is difficult to distinguish good advice from bad one. It is therefore desirable to come up with explainable AI. In this paper, we argue that fuzzy techniques are a proper way to this explainability, and we also analyze which fuzzy techniques are most appropriate for this purpose. Interestingly, it turns out that the answer depends on what problem we are solving: e.g., different “and”- and “or”-operations are preferable when we are controlling a single object and when we are controlling a group of objects.

Kelly Cohen

Department of Aerospace Engineering & Engineering Mechanics, University of Cincinnati,
Cincinnati, OH 45221-0070, USA, e-mail: Kelly.Cohen@uc.edu

Laxman Bokati

Computational Science Program, University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA, e-mail: lbokati@miners.utep.edu

Martine Ceberio

Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA, e-mail: mceberio@utep.edu

Olga Kosheleva

Department of Teacher Education, University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA, e-mail: olgak@utep.edu

Vladik Kreinovich

Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA, e-mail: vladik@utep.edu

1 Why Fuzzy Techniques in Explainable AI

Need for explainable AI. Lately, there have been a new breakthrough in Artificial Intelligence (AI) caused by successes of deep learning; see, e.g., [2].

While deep learning techniques have been very successful in many applications, serious problems surfaced related to their use. The main problem is that a trained neural network is a black box, it does not provide explanations of its decisions.

Since no tool is 100% accurate, it is not clear how to separate correct advice from wrong advice – and in social situations, the advice can be very wrong, repeating the biases that occurred in some parts of the training data.

It is therefore desirable to develop AI tools that would translate numerical recommendations into natural-language explanations.

Why fuzzy in explainable AI. To perform the desired translation between numerical recommendations and natural-language descriptions, it is reasonable to utilize techniques that have been designed in the 1960s – and successfully used since then – specifically to describe the correspondence between natural-language descriptions and numerical recommendations: namely, *fuzzy* techniques; see, e.g., [1, 3, 4, 9, 10, 13]. In these techniques, to each expert statement and to each propositional combination of expert statements, we assign a degree of confidence – a number from the interval $[0, 1]$ describing to what extent we are confident in a given statement – original or combination.

2 Which Fuzzy Techniques in Explainable AI

Challenge. There are many different versions of fuzzy techniques – the main idea is that there are many different “and” and “or”-operations (also known as t-norms and t-conorms), i.e., functions $f_{\&}(a, b)$ and $f_{\vee}(a, b)$ that estimate our degrees of certainty in statements $A \& B$ and $A \vee B$ in situations in which we only know the degrees of confidence a and b in the original statement A and B .

It is known that a wrong choice of an operation can hinder the effectiveness of the resulting system. So which operations should we choose?

Two types of situations. In this paper, we will consider two types of situations:

- In some cases, we are interested in the best performance of an individual system. For example, we have a single drone performing meteorological (or other) observations, and we want to make sure that its probability of failure is as small as possible.
- In other cases, we have a mass phenomenon – e.g., we are controlling a swarm of drones, or a large number of local power stations contributing to the same grid. In this case, we can afford the failure of some of these objects and thus, use less expensive equipment – if this allows us to have more objects and attain the best overall performance.

We will show that in these two types of situations, different pairs of “and”- and “or”-operations are preferable.

Situations where we are interested in the individual performance. In situations when we are interested in the individual performance, in which we want to minimize the probability of failure, we want to deviations of the object from the desired trajectory to be as small as possible – since it is such deviations that cause failure.

One of the possible reasons for such deviations in fuzzy control is that fuzzy control is based on combining the original experts’ degree of confidence by using “and”- and “or”-operations, and the original estimates are only provided with some uncertainty. Just like an expert cannot provide the exact value of the desired control – this is why fuzzy techniques are needed in the first place – the expert also cannot describe his/her degree of confidence in a certain statement by an exact number. If we force the expert to do it – as many systems do – the expert will provide slightly different numbers when asked again about the same statements. These changes affect the results of “and”- and “or”-operations – and thus, affect the resulting control.

A single too-large deviation from the desired control can be disastrous. So, to be on the safe side, we want to make sure that the worst-possible deviation is as small as possible. Let us describe this situation in precise terms. Let $\delta > 0$ denote the accuracy with which the experts can provide their degrees. This means that the same expert can provide estimates a and a' for his/her degree of confidence in the same statement A which are δ -close, i.e., for which $|a - a'| \leq \delta$. Similarly, for another statement B , the expert can provide estimates b and b' for which $|b - b'| \leq \delta$. As a result of this uncertainty, we can have different values $f_{\&}(a, b)$ and $f_{\&}(a', b')$, i.e., we have a non-zero difference $|f_{\&}(a, b) - f_{\&}(a', b')|$.

The worst-case scenario is when this difference is the largest possible. It is characterized by the value

$$w(f_{\&}, \delta) \stackrel{\text{def}}{=} \max_{|a-a'| \leq \delta, |b-b'| \leq \delta} |f_{\&}(a, b) - f_{\&}(a', b')|.$$

We want to select an “and”-operation for which this worst-case value is the smallest possible. Interestingly, it turns out that in this case, the optimal “and”-operation is $f_{\&}(a, b) = \min(a, b)$; see, e.g., [6, 7, 9].

Similarly, for an “or”-operation, the corresponding difference has the form $|f_{\vee}(a, b) - f_{\vee}(a', b')|$. So, the worst-case scenario is when this difference is the largest possible. It is characterized by the value

$$w(f_{\vee}, \delta) \stackrel{\text{def}}{=} \max_{|a-a'| \leq \delta, |b-b'| \leq \delta} |f_{\vee}(a, b) - f_{\vee}(a', b')|.$$

We want to select an “or”-operation for which this worst-case value is the smallest possible. It turns out that in this case, the optimal “or”-operation is $f_{\vee}(a, b) = \max(a, b)$; see, e.g., [6, 7, 9].

So, in situations when we are interested in the individual performance, the optimal selection of fuzzy operations is $f_{\&}(a, b) = \min(a, b)$ and $f_{\vee}(a, b) = \max(a, b)$.

Situations where we are interested in the group performance. In situations where we are interested in the group performance, we allow some systems to fail, but we would like to minimize the number of failing systems – i.e., the probability that a system will fail. A system fails if the corresponding parameters deviate too much from their desired values. Each of these parameters is affected by many different factors. It is known that, under reasonable conditions, the distribution of the joint effect of many independent factors is close to Gaussian; this is known as the Central Limit Theorem; see, e.g., [11].

A normal distribution of each quantity y is uniquely determined by its mean and by its standard deviation. Usually, we can safely assume that the mean is 0 (or close to 0). For a normal distribution with 0 mean and standard deviation σ , the probability of exceeding a threshold value x_0 depends only on the ratio x_0/σ : the larger this ratio – i.e., equivalently, the smaller σ – the smaller this probability.

In general, for a function $y = f(x_1, \dots, x_n)$ of several variables, when the change Δx_i is small, the corresponding change in Δy is approximately equal to $\frac{\partial f}{\partial x_i} \cdot \Delta x_i$, and thus, the corresponding variance σ^2 of y is approximately equal to

$$\left(\frac{\partial f}{\partial x_i} \right)^2 \cdot (\sigma_i)^2,$$

where σ_i is the standard deviation of Δx_i . Thus, to minimize σ , we need to minimize all the values σ_i as well.

In particular, for the result $c = f_{\&}(a, b)$ of an “and”-operation, this means that we need to minimize the standard deviation caused by random deviations Δa and Δb . For small deviations, for each a and b , we have

$$\Delta c = \frac{\partial f_{\&}(a, b)}{\partial a} \cdot \Delta a + \frac{\partial f_{\&}(a, b)}{\partial b} \cdot \Delta b$$

and thus – under the natural assumption that the deviations Δa and Δb are independent and equally distributed, with standard deviation σ_0 – we get

$$\sigma^2(a, b) = \left(\frac{\partial f_{\&}(a, b)}{\partial a} \right)^2 \cdot \sigma_0^2 + \left(\frac{\partial f_{\&}(a, b)}{\partial b} \right)^2 \cdot \sigma_0^2.$$

The overall standard deviation can be obtained by averaging this value over all possible a and b , i.e., by taking

$$\sigma^2 = \int \sigma^2(a, b) da db = \int \left(\left(\frac{\partial f_{\&}(a, b)}{\partial a} \right)^2 \cdot \sigma_0^2 + \left(\frac{\partial f_{\&}(a, b)}{\partial b} \right)^2 \cdot \sigma_0^2 \right) da db.$$

Thus, minimizing the standard deviation means minimizing this integral.

It turns out – see, e.g., [8, 9] – that the “and”-operation $f_{\&}(a, b)$ for which this integral is the smallest possible is $f_{\&}(a, b) = a \cdot b$.

Similarly, for the “or”-operation, we need to minimize the integral

$$\sigma^2 = \int \left(\left(\frac{\partial f_V(a,b)}{\partial a} \right)^2 \cdot \sigma_0^2 + \left(\frac{\partial f_V(a,b)}{\partial b} \right)^2 \cdot \sigma_0^2 \right) da db.$$

It is known (see [8, 9]) that the “or”-operation $f_V(a,b)$ for which this integral is the smallest possible is $f_V(a,b) = a + b - a \cdot b$.

So, in situations when we are interested in the group performance, the optimal selection of fuzzy operations is $f_{\&}(a,b) = a \cdot b$ and $f_V(a,b) = a + b - a \cdot b$.

Other possible situations. We may be instead looking for the operations that lead to the smoothest trajectory, or for the operations that lead to the most stable control, or for the operations which are the fastest to compute. In all these cases, we end up with different pairs of optimal “and”- and “or”-operations; see, e.g., [5, 12].

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), and HRD-1834620 and HRD-2034030 (CAHSI Includes). 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.

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. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
4. J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
5. H. T. Nguyen and V. Kreinovich, “Methodology of fuzzy control: an introduction”, In: H. T. Nguyen and M. Sugeno (eds.), *Fuzzy Systems: Modeling and Control*, Kluwer, Boston, MA, 1998, pp. 19–62.
6. H. T. Nguyen, V. Kreinovich, B. Lea, and D. Tolbert, “How to control if even experts are not sure: robust fuzzy control”. *Proceedings of the Second International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems*, College Station, December 2–4, 1992, pp. 153–162.
7. H. T. Nguyen, V. Kreinovich, and D. Tolbert, “On robustness of fuzzy logics”, *Proceedings of the 1993 IEEE International Conference on Fuzzy Systems FUZZ-IEEE’93*, San Francisco, California, March 1993, Vol. 1, pp. 543–547.
8. H. T. Nguyen, V. Kreinovich, and D. Tolbert, “A measure of average sensitivity for fuzzy logics”, *International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems*, 1994, Vol. 2, No. 4, pp. 361–375.
9. H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.

10. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
11. D. J. Sheskin, *Handbook of Parametric and Non-Parametric Statistical Procedures*, Chapman & Hall/CRC, London, UK, 2011.
12. M. H. Smith and V. Kreinovich. "Optimal strategy of switching reasoning methods in fuzzy control", Chapter 6 in H. T. Nguyen, M. Sugeno, R. Tong, and R. Yager (eds.), *Theoretical aspects of fuzzy control*, J. Wiley, N.Y., 1995, pp. 117–146.
13. L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.