Why Fuzzy Techniques in Explainable AI? Which Fuzzy Techniques in Explainable AI?

Kelly Cohen[1], Laxman Bokati[2], Martine Ceberio[2], Olga Kosheleva[2], and Vladik Kreinovich[2]

[1]Department of Aerospace Engineering & Engineering Mechanics
University of Cincinnati, Cincinnati, OH 45221-0070, USA
Kelly.Cohen@uc.edu
[2]University of Texas at El Paso
500 W. University, El Paso, Texas 79968, USA
lbokati@miners.utep.edu, mceberio@utep.edu,
olgak@utep.edu, vladik@utep.edu

1. Need for Explainable AI

- Lately, there have been a breakthrough in AI caused by successes of deep learning.
- Deep learning techniques have been very successful in many applications.
- However, serious problems surfaced.
- The main problem is that a trained neural network is a black box, it does not provide any explanations.
- Since no tool is 100% accurate, it is not clear how to separate correct advice from wrong advice.
- In social situations, the advice can be very wrong, repeating the biases of the training data.
- It is therefore desirable to develop AI tools that would:
  - translate numerical recommendations
  - into natural-language explanations.

2. Why Fuzzy in Explainable AI

- We want to translate numerical recommendations into natural-language descriptions.
- It is reasonable to utilize *fuzzy* techniques that:
  - have been designed in the 1960s – and successfully used since then,
  - to relate natural-language descriptions and numerical recommendations.
- In these techniques:
  - to each expert statement and to each propositional combination of expert statements,
  - we assign *a degree of confidence*,
  - i.e., a number from the interval $[0, 1]$ describing to what extent we are confident in a given statement.

3. Which Fuzzy Techniques in Explainable AI

- There are many different versions of fuzzy techniques.
- The main idea is that there are many different "and" and "or"-operations.
- These are 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 statements A and B.
- These operations are also known as t-norms and t-conorms.
- It is known that a wrong choice of an operation can hinder the effectiveness of the resulting system.
- So which operations should we choose?

4. Two Types of Situations

- In this talk, 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.
- 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.

5. Two Types of Situations (cont-d)

- 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 "and"- and "or"-operations are preferable.

6. Situations Where We Are Interested in the Individual Performance

- In such situations:
  - we want to minimize the probability of failure,
  - we want the deviations of the object from the desired trajectory to be as small as possible,
  - since it is such deviations that cause failure.
- What are the possible reasons for such deviations in fuzzy control?
- Fuzzy control is based on:
  - combining the original experts' degree of confidence
  - by using "and"- and "or"-operations.

7. Individual Performance (cont-d)

- 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 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.

8. Individual Performance (cont-d)

- 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 for a statement A, the same expert:
  - can provide different estimates a and a' for his/her degree of confidence in A;
  - these estimates are δ-close: $|a - a'| \le \delta$.
- Similarly, for another statement B, the expert can provide estimates b and b' for which $|b - b'| \le \delta$.

9. Individual Performance (cont-d)

- As a result of this uncertainty, we can have different values $f_\&(a,b)$ and $f_\&(a',b')$:

$$|f_\&(a,b) - f_\&(a',b')| \neq 0.$$

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

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

- We want to select an "and"-operation for which this worst-case value is the smallest possible.
- It is known that in this case, the optimal "and"-operation is $f_\&(a,b) = \min(a,b)$.
- Similarly, for an "or"-operation, the corresponding difference has the form $|f_\vee(a,b) - f_\vee(a',b')|$.

10. Individual Performance (cont-d)

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

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

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

$$f_\&(a,b)=\min(a,b) \text{ and } f_\vee(a,b)=\max(a,b).$$

11. Situations Where We Are Interested in the Group Performance

- In such situations, we may allow some systems to fail.
- 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.

12. Group Performance (cont-d)

- 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.

13. Group Performance (cont-d)

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

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

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

14. Group Performance (cont-d)

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

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

- The natural assumption is that the deviations Δa and Δb are i.i.d., with standard deviation σ_0.
- Then, we get

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

15. Group Performance (cont-d)

- The overall standard deviation can be obtained by averaging this value over all possible a and b:

$$\sigma^2 = \int \sigma^2(a,b)\,da\,db =$$

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

- Thus, minimizing the standard deviation means minimizing this integral.
- It turns out that the "and"-operation $f_\&(a,b)$ for which this integral is the smallest possible is $f_\&(a,b) = a \cdot b$.

16. Group Performance (cont-d)

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

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

- It is known that the "or"-operation $f_\vee(a,b)$ for which this integral is the smallest possible is

$$f_\vee(a,b) = a + b - a \cdot b.$$

- So, in situations when we are interested in the group performance, the optimal fuzzy operations are:

$$f_\&(a,b) = a \cdot b \text{ and } f_\vee(a,b) = a + b - a \cdot b.$$

17. Other Possible Situations

- We may be looking for the operations:
  - that lead to the smoothest trajectory, or
  - that lead to the most stable control, or
  - which are the fastest to compute.
- In all these cases, we end up with different pairs of optimal "and"- and "or"-operations.

18. 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).

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.