Foundations of Explainable Fuzzy AI *and* Interval Computations: *An Overview of Forthcoming Classes*

Vladik Kreinovich

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

1. Classes

- Foundations of Explainable Fuzzy AI
  CS 5454/CS 4365
  Summer II class MTWRF 11:40-1:50 pm
- Interval Computations
  CS 5351/CS 4365
  Fall 2021 class TR 3-4:20 pm

Part I
Explainable Fuzzy AI Class

2. Need for Explainable AI

- Modern AI techniques – especially deep learning – provide, in many cases, very good recommendations:
  - where a self-driving car should go,
  - whether to give a company a loan, etc.
- The problem is that these techniques are not (yet) perfect.
- In some cases, the recommendations generate by an AI system are not good.
- Of course, as the famous Marilyn Monroe movie says, "Nobody's perfect".
- Human experts are not perfect either; however:
  - when a human expert – be it a banking official or a medical doctor – makes a recommendation,
  - he or she can, if asked, provide an explanation.

3. Need for Explainable AI (cont-d)

- If you find the explanation not sufficiently convincing, you can ask for someone else's advice.
- Unfortunately:
  - recommendations provided by an AI system (such as a deep neural network)
  - usually come without an explanation.
- So we cannot so easily separate good and bad advice.
- It is therefore desirable to make AI more explainable.

4. Why Fuzzy Techniques

- Providing an explanation means:
  - finding natural-language rules and ideas which are,
  - in some reasonable sense, equivalent to the numerical results provided by the AI tools.
- The problem of connecting natural language rules and numerical decisions is known since 1960s.
- Then, the need was recognized to incorporate expert knowledge into control and decision making.
- Expert use imprecise words like "small".
- For this incorporation, a special technique was invented – known as fuzzy techniques.
- This technique led to many successful applications.
- It is therefore reasonable to use this techniques in designing explainable AI.

5. What We Will Study in This Class

- If we knew how to make AI explainable, teaching this class would be easier.
- We would just teach the corresponding algorithms and methods.
- At present, explainable AI remains largely an ultimate goal.
- We do not yet know which tools will work better.
- So, instead of studying specific tools, it makes sense to study the *foundations* for these tools, so that:
  - we will know why we need to use these tools, and
  - we will know which tools are better in which situations.
- This will help us select appropriate tools for making current AI applications more explainable.

6. First Topic: Introduction to Fuzzy Techniques

- We want to better understand how fuzzy techniques can help with explainable AI.
- For this, we need to have a good understanding of these techniques.
- We will learn the corresponding techniques and how they are used in control and in other applications.
- We will also try to make these techniques themselves more explainable.
- Namely, we will explain first-principle motivations for these techniques.

7. Introduction to Fuzzy Techniques (cont-d)

- We will study all three main stages of fuzzy techniques:
  1. describing the original imprecise words like "small" in numerical terms,
  2. combining the corresponding numbers:
     - to describe boolean (and- and or-) combinations of the corresponding properties
     - special "and" and "or" operations are used for this;
  3. "defuzzification" – transforming imprecise recommendations into a precise control value.

8. Second Topic: Which Version of Fuzzy Technique to Select

- In all three stages of fuzzy techniques, there are several different options.
- Empirically, in different situations, different options work best.
- This makes sense, since in different situations, we have different objectives.
- For example:
  - if launch a single drone to inspect an area,
  - the main objective is to maximize the probability that its mission succeeds.

9. Which Version of Fuzzy (cont-d)

- On the other hand:
  - if we launch a swarm of drones to inspect the same area,
  - it is probably Ok if one of them does not do much,
  - as long as, on average, the overall mission is successful.
- How do we select the best techniques?
- In some cases, we have finitely many parameters.
- So we need to find the best values of these parameters.
- To find the largest and the smallest values of a function of several such variables, we can use calculus.
- Do not worry if you have forgotten some of it, we will refresh.

10. Which Version of Fuzzy (cont-d)

- In many other cases, however, we need to select a function – e.g., the best "and"- and "or"-operations.
- There is a natural generalization of calculus that deals with such optimization problems.
- It is known as *variational calculus*, and it is actively used in control.
- We will learn the basics of this techniques.
- As an example, we will use this technique:
  - to come up with optimal "and"- and "or"-operations
  - for the two above-described drone situations.

11. Third Topic: Towards Explainable Machine Learning

- The ultimate goal is to make the *results* of machine learning (and other AI techniques) explainable.
- We are still working on this.
- Meanwhile, an important help would be to make the machine learning techniques themselves explainable.
- At present, in many cases:
  - the only reason we select some techniques and some parameters of these technique
  - is that these techniques empirically work well on several problems.
- This is not as convincing as when we prove that these techniques are, in some reasonable sense, optimal.
- We will analyze deep learning from this viewpoint.

12. Explainable Machine Learning (cont-d)

- We will show that many empirically successful features of deep learning can indeed be proven to be optimal.
- The corresponding proofs require the use of:
  - other techniques widely used in applications to the physical world,
  - namely, the technique of symmetries.
- So, we will learn symmetry-related optimization techniques as well.

13. Projects

- This course is somewhat on the theoretical side.
- But it is important to have applications in mind.
- Students will be therefore encouraged to work on projects.
- Ideally, projects should be related to real-life applications.
- Purely theoretical projects are also OK.
- A project may consist of reviewing some paper(s):
  - on explainable AI,
  - on its applications, and
  - on its foundations.

14. Projects (cont-d)

- It is also possible to select a more creative project.
- In such a project, students – individually or in groups – will come up with something new.
- Explainable AI is a new developing topic, there are many more open problems than results.
- Any progress in any of these open problems will bring us closer to the goal of making AI explainable.

Part II
Interval Computations Class

15. Need for Indirect Measurements

- We are often interested in a quantity y that is difficult or even impossible to measure directly.
- To estimate the value of y, we:
  - measure easier-to-measure quantities x_i which are related to y by a known dependence
$$y = f(x_1, \ldots, x_n), \text{ and}$$
  - use the measured values $\widetilde{x}_i$ of these quantities to estimate y as $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$.
- In engineering, this procedure is called *indirect measurement.*
- In computer science, it is called *data processing.*
- The need for extensive data processing is the main reason why computers were invented in the first place.

16. Measurements Are Never 100% Exact

- Problem: measurements are never 100% exact.
- For example, a person is not exactly 5 ft 7 in, the actual height is slightly different.
- Because of the differences $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i \neq 0$, the result $\widetilde{y}$ of data processing differs from the actual value y:

$$\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) \neq y = f(x_1, \ldots, x_n).$$

- Traditionally, computers simply return a number $\widetilde{y}$.
- They do not specify how accurate is this number, i.e., how big the inaccuracy $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$ can be.

17. Measurements Are Not 100% Exact (cont-d)

- From the practical viewpoint, however, it is important to know this accuracy.
- For example, if we estimated the amount of oil in a field as 100 million ton, then:
  - if it is 100 ± 10, this is good news, we should start exploiting this area;
  - however, if it is 100 ± 200, then maybe there is no oil at all,
  - so it would be better to perform additional measurements before investing into this area.

18. Traditional Approach Assumes That We Know Probabilities

- Traditional engineering approach to this uncertainty is what students learn, e.g., in engineering labs.
- It assumes that we know the probabilities of different values of measurement error Δx_i.
- Usually, it is assumed that there errors are independent and normally distributed.
- In practice, often, we do not know these probabilities.
- In many practical situations, we only know the upper bounds Δ_i on the measurement errors: $|\Delta x_i| \le \Delta_i$.
- This means that:
  - after we measure the value as $\widetilde{x}_i$,
  - the actual value of the measured quantity can be anywhere within the interval $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.

19. Sometimes, We Do Not Know Probabilities

- For example, if a thermometer shows 86 F, and its accuracy is half a degree:
  - it does not mean that the temperature is exactly 86.000,
  - it could be anywhere between 85.5 and 86.5.
- For a single measurement result, this inaccuracy is easy to handle.
- However, when we process thousands of data points, these inaccuracies add up.

20. Main Problem of Interval Computations

- Sometimes, the only thing we know about each input x_i is the interval $\mathbf{x}_i$ of possible values.
- Then the natural question is: what is the range $\mathbf{y}$ of possible values of $f(x_1, \ldots, x_n)$?
- The problem of computing this range is called the problem of *interval computations*.

21. Where Are Interval Computations Used?

- They started with DoD making sure that the intercontinental missiles fall within the desired range.
- They continued with NASA designing a trajectory which is guaranteed to hit the Moon.
- Nowadays, they are used everywhere:
  - in manufacturing,
  - in robotics,
  - in geoinformatics (processing geophysical data),
  - in bioinformatics (processing bioinformatics data),
  - in economics,
  - in computer graphics,
  - in computer security and privacy (where intervals are introduced to avoid disclosing exact values).

22. What Will Be in the Course

- Formulation of the main problem.
- General (software) methods of solving this problem.
- In general, the problem of finding the range $\mathbf{y}$ is NP-hard (intractable).
- Therefore, we can expect only estimates for $\mathbf{y}$.
- For each estimation method, we will describe:
  - the requited computation time,
  - the possibility of parallelization, etc.
- We will also have a brief survey of applications.

23. Projects

- After learning the basic interval techniques, students will start working on an (individual or group) project.
- Ideally, a project should be:
  - related to the main topic of student's research (we can help with that) and
  - help the student in working on his/her dissertation, thesis, or project.
- Some students have not yet selected a topic.
- For them, here are many interesting open research problems in which they can help.